

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 15:49:44 ; Search time 1021.63 Seconds

(Without alignments)
216.534 Million cell updates/sec

Title: US-09-593-914-1

Sequence: 1 agcgggtctataga 15

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_bal:*
2: gb_bal:*
3: gb_bal:*
4: gb_bal:*
5: gb_bal:*
6: gb_bal:*
7: gb_bal:*
8: gb_bal:*
9: gb_bal:*
10: gb_bal:*
11: gb_bal:*
12: gb_bal:*
13: gb_bal:*
14: gb_bal:*
15: gb_bal:*
16: gb_bal:*
17: gb_bal:*
18: gb_bal:*
19: gb_bal:*
20: gb_bal:*
21: gb_bal:*
22: gb_bal:*
23: gb_bal:*
24: gb_bal:*
25: gb_bal:*
26: gb_bal:*
27: gb_bal:*
28: gb_bal:*
29: gb_bal:*
30: gb_bal:*
31: gb_bal:*
32: gb_bal:*
33: gb_bal:*
34: gb_bal:*
35: gb_bal:*
36: gb_bal:*
37: gb_bal:*
38: gb_bal:*
39: gb_bal:*
40: gb_bal:*
41: gb_bal:*
42: gb_bal:*
43: gb_bal:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_vl3:*
61: gb_vl4:*
62: gb_vl5:*
63: gb_vl6:*
64: gb_vl7:*
65: gb_vl8:*
66: gb_vl9:*
67: gb_vl10:*
68: gb_vl11:*
69: gb_vl12:*
70: gb_vl13:*
71: gb_vl14:*
72: gb_vl15:*
73: gb_vl16:*
74: gb_vl17:*
75: gb_vl18:*
76: gb_vl19:*
77: gb_vl20:*
78: gb_vl21:*
79: gb_vl22:*
80: gb_vl23:*
81: gb_vl24:*
82: gb_vl25:*
83: gb_vl26:*
84: gb_vl27:*
85: gb_vl28:*
86: gb_vl29:*
87: gb_vl30:*
88: gb_vl31:*
89: gb_vl32:*
90: gb_vl33:*
91: gb_vl34:*
92: gb_vl35:*
93: gb_vl36:*
94: gb_vl37:*
95: gb_vl38:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	9 AX057545	AX057545 Sequence
2	15	100.0	1787	13 BB18RNA	BB18RNA
3	15	100.0	1799	13 BB18RNA	BB18RNA
4	15	100.0	1799	13 BB18RNA	BB18RNA
5	15	100.0	1799	13 BB18RNA	BB18RNA
6	15	100.0	198038	65 AC018980	AC018980 Homo sapi
7	15	100.0	209213	66 AC020644	AC020644 Homo sapi
8	14	93.3	39547	64 AC016460	AC016460 Homo sapi
9	14	93.3	45300	12 AC007288	AC007288 Arabidops
10	14	93.3	55739	13 AP000389	AP000389 Arabidops
11	14	93.3	81414	12 AB024037	AB024037 Arabidops
12	14	93.3	110157	12 AF058825	AF058825 Arabidops

```

12      14      93.3 123080 14 F10A2
13      14      93.3 170390 88 AC026473
14      14      93.3 175462 87 AC009169
15      14      93.3 175742 81 AL391534
16      14      93.3 183181 13 AUCR1919
17      14      93.3 184427 58 EH020824
18      14      93.3 184869 64 AC015622
19      14      93.3 211999 77 AC087039
20      14      89.3 705 14 HU34210
21      13.4 89.3 705 14 HU34210
22      13.4 89.3 705 14 HU34213
23      13.4 89.3 705 14 HU34221
24      13.4 89.3 706 14 IB04209
25      13.4 89.3 706 14 IB04209
26      13.4 89.3 706 14 IB04209
27      13.4 89.3 706 14 IB04209
28      13.4 89.3 706 14 IB04209
29      13.4 89.3 706 14 IB04209
30      13.4 89.3 706 14 IB04209
31      13.4 89.3 706 14 IB04209
32      13.4 89.3 706 14 IB04209
33      13.4 89.3 706 14 IB04209
34      13.4 89.3 706 14 IB04209
35      13.4 89.3 706 14 IB04209
36      13.4 89.3 706 14 IB04209
37      13.4 89.3 706 14 IB04209
38      13.4 89.3 706 14 IB04209
39      13.4 89.3 706 14 IB04209
40      13.4 89.3 706 14 IB04209
41      13.4 89.3 706 14 IB04209
42      13.4 89.3 706 14 IB04209
43      13.4 89.3 706 14 IB04209
44      13.4 89.3 706 14 IB04209
45      13.4 89.3 706 14 IB04209

```

ALIGNMENTS

```

RESULT 1
LOCUS AX057545 15 bp DNA PAT 17-JAN-2001
DEFINITION Sequence 1 from Patent WO0072759.
ACCESSION AX057545
VERSION AX057545.1 GI:12310273
KEYWORDS
SOURCE Dekkera bruxellensis.
ORGANISM Dekkera bruxellensis.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetales;
Saccharomycetaceae; Dekkera.

```

```

REFERENCE 1
AUTHORS Hyding-Nielsen, J.J., O'Keefe, H.P., and Stender, H.
TITLE Probes, probe sets, methods and kits pertaining to the detection,
JOURNAL Identification and/or enumeration of yeast; particularly in wine
Boston Probes, Inc. (US)

```

```

FEATURES
Source location/Qualifiers
1..15
/organism="Dekkera bruxellensis"
/db_xref="taxon:5007"
/note="Description of Combined DNA/RNA molecule: PROBING
Nucleotide Sequence"

```

```

BASE COUNT 4 a 2 c 5 g 4 t
ORIGIN

```

```

Query Match 100.0%; Score 15; DB 9; Length 15;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 agcggtctataga 15
Db 1 AGCGGCTCTATAGA 15

```

```

RESULT 2
LOCUS B18RNA/c 1787 bp DNA PLN 17-SEP-1998
DEFINITION B. bruxellensis 18S rRNA gene (NCYC 370).
ACCESSION X83815
VERSION X83815.1 GI:1913922
KEYWORDS 18S ribosomal RNA.
SOURCE Brettanomyces bruxellensis.
ORGANISM Brettanomyces bruxellensis.
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Candidaceae; Brettanomyces.

```

```

REFERENCE 1
AUTHORS Cai, J., and Collins, M.D.
JOURNAL Unpublished
2 (bases 1 to 1787)

```

```

REFERENCE 3
AUTHORS Cai, J., Roberts, I.N., and Collins, M.D.
JOURNAL Submitted (11-JAN-1995) J. Cai, BBSRC, Institute of Food Research,
Reading Lab, Earley Gate, Whiteknights Road, Reading RG6 2EP, UK
3 (bases 1 to 1787)

```

```

REFERENCE 1
AUTHORS Cai, J., Roberts, I.N., and Collins, M.D.
JOURNAL Phylogenetic relationships among members of the ascomycetous yeast
genera Brettanomyces, Debaryomyces, Dekkera, and Kluyveromyces
deduced by small-subunit rRNA gene sequences
Int. J. Syst. Bacteriol. 46 (2), 542-549 (1996)
97089000

```

```

FEATURES
Source location/Qualifiers
1..1787
/organism="Brettanomyces bruxellensis"
/db_xref="taxon:37784"
1..1787
/product="18S ribosomal RNA"

```

```

BASE COUNT 471 a 353 c 464 g 469 t 10 others
ORIGIN

```

```

Query Match 100.0%; Score 15; DB 13; Length 1787;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 agcggtctataga 15
Db 1053 AGCGGCTCTATAGA 1039

```

```

RESULT 3
LOCUS B18RNA/c 1799 bp DNA PLN 17-SEP-1998
DEFINITION B. bruxellensis 18S rRNA gene (NCYC 362).
ACCESSION X84814
VERSION X84814.1 GI:1913923
KEYWORDS 18S ribosomal RNA.
SOURCE Brettanomyces bruxellensis.
ORGANISM Brettanomyces bruxellensis.
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Candidaceae; Brettanomyces.

```

```

REFERENCE 1
AUTHORS Cai, J., and Collins, M.D.
JOURNAL Unpublished
2 (bases 1 to 1799)

```

```

REFERENCE 3
AUTHORS Cai, J., Roberts, I.N., and Collins, M.D.
JOURNAL Direct Submission
Submitted (11-JAN-1995) J. Cai, BBSRC, Institute of Food Research,
Reading Lab, Earley Gate, Whiteknights Road, Reading RG6 2EP, UK
3 (bases 1 to 1799)

```

```

REFERENCE 1
AUTHORS Cai, J., Roberts, I.N., and Collins, M.D.
JOURNAL Phylogenetic relationships among members of the ascomycetous yeast
genera Brettanomyces, Debaryomyces, Dekkera, and Kluyveromyces
deduced by small-subunit rRNA gene sequences
Int. J. Syst. Bacteriol. 46 (2), 542-549 (1996)

```

MEDLINE 97089000 Location/Qualifiers
 FEATURES
 source 1..1799
 /organism="Brettanomyces bruxellensis"
 /strain="MCYC 362 non type strain"
 /db_xref="taxon:37784"
 rRNA 1..1799
 /product="18S ribosomal RNA"
 /partial
 BASE COUNT 476 a 358 c 487 g 475 t 3 others
 ORIGIN

Query Match 100.0%; Score 15; DH 13; Length 1799;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 aagagctctattaga 15
 |||||
 Db 1065 ACCCGCTATTAGA 1051

RESULT 4
 HB18SRKNA/c HB18SRKNA 1799 bp DNA PLN 08-DEC-1992
 LOCUS D.bruxellensis gene for 18S ribosomal RNA.
 DEFINITION X58052
 ACCESSION X58052.1 GI:2485
 VERSION 18S ribosomal RNA.
 KEYWORDS Dekkera bruxellensis.
 SOURCE Dekkera bruxellensis.
 ORGANISM Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Dekkera.

REFERENCE 1 (bases 1 to 1799)
 De Wachter R.
 TITLE Direct Submission
 JOURNAL Submitted (22-FEB-1991) R. De Wachter, Universiteit Antwerpen (UIA), Universiteitsplein 1, 2610 Wilrijk, Belgium
 REFERENCE 2 (bases 1 to 1799)
 Hendriks L., Goris A., Van de Peer Y., Neels J.M., Vancanneyl M., Kersters K., Berty J.F., Hennebert G. and De Wachter R.
 TITLE The evolution of ascomycetes and ascomycete-like yeasts as inferred from small ribosomal subunit RNA sequences
 JOURNAL Syst. Appl. Microbiol. 15, 98-104 (1992)

FEATURES
 source 1..1799
 /organism="Dekkera bruxellensis"
 /strain="MCYC 27700"
 /db_xref="taxon:5007"
 /genome
 rRNA <1..>1799
 /gene="18S rRNA"
 /product="18S ribosomal RNA"
 1..1799
 /gene="18S rRNA"
 BASE COUNT 479 a 359 c 483 g 478 t
 ORIGIN

Query Match 100.0%; Score 15; DH 13; Length 1799;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 aagagctctattaga 15
 |||||
 Db 1066 ACCCGCTATTAGA 1052

RESULT 5
 AC018980/c AC018980 198038 bp DNA HTG 24-AUG-2000
 LOCUS Homo sapiens chromosome 10 clone RP11-476F14, WORKING DRAFT
 DEFINITION
 SOURCE, 9 unordered pieces.

ACCESSION AC018980
 VERSION AC018980.5 GI:9887658
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 198038)
 Smith, D.K.
 TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome Sequence Data Unpublished
 2 (bases 1 to 198038)
 Smith, D.K.
 JOURNAL Direct Submission
 REFERENCE Submitted (25-DEC-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
 COMMENT On Aug 24, 2000 this sequence version replaced gi:8389444.
 Genome Center
 Center: Genome Therapeutics Corporation
 Center code: GTC
 Web site: <http://www.gnomecorp.com/>
 Contact: gtc-sequencer@genomecorp.com
 Project Information
 Center project name: hg081
 Summary Statistics
 Sequencing vector: N/A
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 990315
 Consensus quality: 181125 bases at least Q40
 Consensus quality: 191464 bases at least Q30
 Consensus quality: 193045 bases at least Q20
 Insert size: 197298; sum-of-contrigs
 Quality coverage: 5.1x in Q20 bases; sum-of-contrigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contrigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contrigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 3231: contrig of 3231 bp in length
 * 3332 3332: gap of unknown length
 * 11374 11374: contrig of 8043 bp in length
 * 11475 11475: gap of unknown length
 * 25209 25209: contrig of 13735 bp in length
 * 25210 25210: gap of unknown length
 * 45410 45410: contrig of 6376 bp in length
 * 31686 31686: gap of unknown length
 * 31786 31786: gap of unknown length
 * 42742 42742: contrig of 10956 bp in length
 * 42842 42842: gap of unknown length
 * 59298 59298: contrig of 16457 bp in length
 * 59398 59398: gap of unknown length
 * 59399 59399: contrig of 30483 bp in length
 * 89881 89881: gap of unknown length
 * 89882 89882: contrig of 32560 bp in length
 * 122341 122341: gap of unknown length
 * 122442 122442: gap of unknown length
 * 122442 122442: contrig of 75597 bp in length.

FEATURES
 source 1..198038
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-476F14"
 /clone_id="RP11-476F14"
 1..3231
 /note="assembly_name:Contig23"
 3332..11374
 /note="assembly_name:Contig24"
 11475..25209
 /note="assembly_name:Contig25"
 misc_feature
 misc_feature
 misc_feature

*	25214	26025:	conf	of	811	bp	in	length
*	26026	26125:	gap	of	100	bp		
*	26126	26917:	conf	of	792	bp	in	length
*	26918	27017:	gap	of	100	bp		
*	27018	27819:	conf	of	802	bp	in	length
*	27820	27919:	gap	of	100	bp		
*	27920	28727:	conf	of	808	bp	in	length
*	28728	28827:	gap	of	100	bp		
*	28828	29610:	conf	of	783	bp	in	length
*	29611	29710:	gap	of	100	bp		
*	29711	30552:	conf	of	813	bp	in	length
*	30552	30625:	gap	of	100	bp		
*	30624	31419:	conf	of	796	bp	in	length
*	31420	31510:	gap	of	100	bp		
*	31520	32324:	conf	of	805	bp	in	length
*	32325	32421:	gap	of	100	bp		
*	32425	33213:	conf	of	789	bp	in	length
*	33214	33313:	gap	of	100	bp		
*	33314	34121:	conf	of	808	bp	in	length
*	34122	34221:	gap	of	100	bp		
*	34222	35025:	conf	of	804	bp	in	length
*	35026	35125:	gap	of	100	bp		
*	35126	35947:	conf	of	822	bp	in	length
*	35948	36047:	gap	of	100	bp		
*	36048	36848:	conf	of	801	bp	in	length
*	36849	36948:	gap	of	100	bp		
*	36949	37741:	conf	of	793	bp	in	length
*	37742	37841:	gap	of	100	bp		
*	37842	38625:	conf	of	784	bp	in	length
*	38626	38725:	gap	of	100	bp		
*	38726	39547:	conf	of	822	bp	in	length

	a	c	q	t.	others
BASE COUNT ORIGIN	10768	7134	7055	10151	4439

Query Match 93.3%; Score 14; DB 64; Length 39547;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```
QY      1 agcgggtctatag 14
          |||||
Db 10427 AGCGGCTCTATAG 10440
```

RESULT	8		
AC007288			
LOCUS	AC007288	45200 bp	13A
DEFINITION	Arabidopsis thaliana chromosome II section 69 of 255 of the complete sequence. Sequence from clones F10C8.		
ACCESSION	AC007288	AE002093	

ORGANISM
Arabioidae: *Thaliana*
Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons: core eudicots; Rosidae: eurosids II;
Brassicales: Brassicaceae; Arabidopsis.
1 (bases 1 to 4530)
REFERENCE
1. In: X., Kaut, S., Nounisley, S.D., Shea, N.P., Henito, M.-I., Town, C.D.,
AUTHORS

TITLE	Author
Full-length cDNA sequence of the <i>Arabidopsis thaliana</i> <i>At1g19800</i> gene	Fujiki, C.Y., Mason, T.M., Bowman, C.L., Barnslee, M.E., Feldhlyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.Y., Ronting, C.M., Koo, H., Moffit, R.S., Cronin, L.A., Shen, M., Vankken, S.E., Umayah, L., Rallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Ciesky, T.H., Goodman, H.M., Somerville, C.R., Copenhagen, G.P., Preuss, D., Netman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M., and Venter, J.C.
Sequence and analysis of chromosome 2 of the plant <i>Arabidopsis thaliana</i>	

Journal: *thaliana*
 Nature 402 (6763), 761-768 (1999)
 MEDLINE: 20083487
 PUBMED: 10617197
 REFERENCE: 2 (bases 1 to 45300)
 AUTHORS: Lin, X.
 TITLE: Direct Submission
 JOURNAL: Submitted (09-MAR-2000) The Institute for Genomic Research, 9712

Genes were identified by a combination of three methods: Gene prediction programs including GAIL (<http://arthur.washington.edu/pub/gail/>), GeneLinder (Phil Green, University of Washington), GenScan (Chris Burge, <http://genome.stanford.edu/GENSCAN.html>), and N-Plantane (<http://www.cbs.dtu.dk/services/NetGene2/>). Searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curation based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/edu/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/MX/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/Westhill/ABI consortium for sequencing BAC clones F6P23, F516, T17A5, and T13L16; the ESSA group for sequencing clone R1304, and Scott Jackson, Jiating Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Microinformatics Department, especiallyixin Zhou, Haniul Kihlak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Belwood Richardson for software and database support.

This work was supported by the National Science Foundation Department of Energy and the US Department of Agriculture.

Address all correspondence to: atm@igp.org
www.igp.org

```

Source
1. 45300
/organism "Arabidopsis thaliana"
/contigvar "Columbia"
/dh_xref-"taxon:3702"
/chromosome-"11"
/complement(1..>45300)
/note="Sequence from clone F10C8"
complement(98..128)
/rpl_family="(TAAA)n"
1561..7051
/gene="At2g12650"
/note="F10C8.11"
/pseudo
4389..4765
/note="DNA sequence related to non-LTR TAI1 family of
retrotransposons"
9531..9641
/rpl_family="(TAAA)n"
complement(10987..11581)
/gene="At2g12660"
/note="F10C8.10"
/pseudo
12276..15124
gene

```

```

/gene="At2g12670"
/note="p10c8.9"
/pseudo
repeat_region
12647..12802
/note="DNA sequence related to non-LTR TAI family of
retrotransposons"
repeat_region
complement(19134..19261)
/rpt_family="(TA)n"
gene
21297..25308
/gene="At2g12680"
/note="p10c8.8"
/pseudo
repeat_region
complement(22159..22261)
/rpt_family="(CAAA)n"
repeat_region
22681..22732
/rpt_family="(GAA)n"
repeat_region
complement(26355..26475)
/rpt_family="(AAAA)n"
gene
complement(26838..27372)
/gene="At2g12690"
/note="p10c8.7"
/pseudo
repeat_region
complement(27694..35519)
/note="this region contains 747 bp tandem repeat units,
like the one in bp 23172..23914. There is a 20 kb gap
between 27227 and 27286. The entire region (27 kb) thus
contains about 37 copies of the repeat units."
misc_feature
complement(31405..31464)
/note="gap of about 20 kb. This gap contains pure tandem
repeats of 747 bp. We are currently trying different
strategies to try to close this gap."
mRNA
complement(join(36774..36993,37088..37323))
/gene="At2g12700"
complement(<36774..>37323)
/gene="At2g12700"
/note="p10c8.6: similar to GH:AD23022"
CDS
complement(join(36774..36993,37088..37323))
/gene="At2g12700"
/note="hypothetical protein"
/codon_start=1
/protein_id="A031080.1"
/db_xref="GI:4850413"
/translation="MNPIDGLPLRDVPAVLTAPRADIPYLPNGICITWYIKITVTWTF
KFNHOLAATICDILQNRHHAMITTYRMIPVYKASTICMHTWDSRHTIIVWQFINVA
KTLRDNRAKRAKRAAGDLSKRMATTIVMDLYDMFTTEGPDVGFMS"
gene
37987..39001
/gene="At2g12710"
/note="p10c8.5"
/pseudo
CDS
complement(join(39615..40076,40213..41085,41290..42414))
/gene="At2g12720"
complement(<39615..>42414)
/gene="At2g12720"
/note="p10c8.4: similar to MUKA transposase of maize
Mutator transposon"
CDS
complement(join(39615..40076,40213..41085,41290..42414))
/gene="At2g12720"
/codon_start=1
/product="Mutator-like transposase"
/protein_id="A031079.1"
/db_xref="GI:4850412"
/translation="MAGPATVIVVSGSWKKORFVNPDERCGRVQLAEKAVYDSL
LKMVDDYGVQNSTHQRSLSYFMSKTKLKNPEFTPPVYTSNRQLCYLITKEQL
RLCVFTAKDSSDISRKDGSRGSKRPLESDSRAESRFEPWGEVGYCYQDEDA
GDERGDGSEVDEADSEEDDEDESRPFDDSGASSDEDEAVYGEPRKN
DEDSKPSKRIQJMEMTOSAGSLTDLKLSTLLAVCOOYRSKFELEYKLLAI
RDGEFDPVTSNKTIVSYKQWDRCLMRACRQGNPNFYVYIYNSHETGSRRCG
RSROATPVIVGLVRYDLDGVDYKPKSICITITKPRKASTSYSTLFRARELT
IGTDSSEELPSYLVIRANPGVYALQDDEGRNMFYVFGASIGFHYRRELV
VYDGTTHGSKGILLALADGNGFQIFPLAFGVNTDENSRLWLTOLKAVYDAR
DUATISDHSISGALGEVPLAARGICTYLLYLNLYFKRKDLPLVKKAYCYRL
NDPTNAENIEEDPLILHAYLQAGCEWMAHAFNPGDNYNLTNTNAESMNAISQAK
NLPVIRMEIAIROMMTWFEERRHDSKQHTQLPQVEKILLQTRVYSSNLIDVQIDA

```

```

repeat_region
BASIC COUNT 1414 a 7779 c 8978 g 14339 t 60 others
ORIGIN
complement(41851..41948)
/rpt_family="(GAA)n"

```

```

Query Match 93.3%; Score 14; DB 12; Length 45500;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Ox 2 gccggtcattaga 15
|||||
Db 40071 GCCGCTATTAGA 40084

```

```

RESULT 9
LOCUS AP000389 55739 bp DNA PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, pl clone:MSJ3.
ACCESSION AP000389 BA000014
VERSION AP000389.1 GI:5672589
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
Kaneko, T., Katoh, T., Sato, S., Nakamura, A., Asamizu, E. and Tabata, S.
Structural analysis of Arabidopsis thaliana chromosome 3. II.
Sequence features of the 4,251,695 bp regions covered by 90 pl, TAC
and BAC clones
DNA Res. 7 (3), 217-221 (2000)
2 (bases 1 to 55739)
Kaneko, T., Katoh, T., Sato, S., Nakamura, Y., Asamizu, E. and Tabata, S.
Direct Submission
Submitted (28-JUL-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail: ynakam@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/sgd_graph.cgi?c=MSJ3
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Graal
(International Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Graal-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
McKenzie (S.M. Heysaard, et al., CDS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/McKenzie/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremlii.zool.lastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MVAL1 and the 3' clone is MOP15.
Location/Qualifiers
1..55739
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MSJ3"
/clone_1fb="Mitsui P1"

```

FEATURES

source

CDS
join(462..735,1345..1373)
/note="gb|AAC61810.1
gene_id:MSJ3.1
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01420.1"
/db_xref="GI:11994245"
/translation="MDPKAEISRLVLPDILGEIYAKIGAKYADYINCILSKEL
GASADEKVLKTLNLAFLVKPLSCSKKILLKCKLANNNPDHAYIKANSRNRK"
complement(join(2420..4168,4328..5023,5138..5459))
/note="contains similarity to Arabidopsis thaliana
retrotransposon Athila
gene_id:MSJ3.2"
/codon_start=1
/pseudo
/evidence-not_experimental
6784..8561
/note="gene_id:MSJ3.3
unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
8666..9013
/note="gb|AAP24530.1
gene_id:MSJ3.4
similar to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
join(13427..13562,13605..13769,14564..15219,15302..15409,
15500..15556,15681..15725,15808..16049,16127..16235,
16688..16859,16997..17124,17227..17382)
/note="gb|AAC97213.1
gene_id:MSJ3.5
strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
join(19136..20065,20144..20368,20467..20550,20807..20926,
21101..21220,21809..21955,22036..22182,22263..22353,
22432..22550,22624..22794,22894..22941)
/note="gb|AAD25853.1
gene_id:MSJ3.6
strong similarity to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01421.1"
/db_xref="GI:11994246"
/translation="MSIKLIPFKRLPVGSEVNIIRVSKHSSEKIVSVKALIEPI
EFTLMDFLPVFEADLAFSSHIVHLLIQRRLTNNDELMEVADQPKRSLE
EITTLGEMDASSSSVNRKREKRWIKUNOTLKSKQVWIKREMTFILTYVL
EKRSKMPERELRSLVEGLILANRPTMLTLENKRAIDEPKPKYFMWQIVY
SVLVHVRKIGNDYILRKQYGLPGVDSIOIWAISVOKICGHGCRDSSSGNPII
LQMSSTSPSAKATIMLADQIKVHPLSCGFYVLOKRLFTTDRKRLIKQILDPVIVY
KQRSKSKDWGCMIOVROGHOKIDHVAIKVISTVHOSRRKRSPOKRRKRRG
RKQDEEKKOEEGKEELEKVEYRGDEGTEKQEIPOQDEMEGSEKQDEEGEE
EKVEYRGDEGTEKQEIPOQDEMEGSEKQDEEGKEEVEKVLKEEVEEHEE
TEDEEAVVILSDDEDGATAPTESQPKETTEVEKEEVEEHEDETEDEEAVV
LSDDEDGATAPTESQPKETTEVEKEEVEEHEDETEDEEAVVILSDDEDGAT
PLTPVMIQSEVMEIDLVKWKAKKKLIRLLSLFHLLIMPSSKQKYLINMIDK
YKIAKKATILLIPDKILAAHRYVYIVQICLIINVKDNATKAKMDIVL"
complement(join(23766..24009,24128..25699,25824..25874,
26025..26075,26157..26453))
/note="gene_id:MSJ3.7"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="Mutator-like transposase"
33238..33464
/note="contains similarity to En/Spm-like transposon
protein
gene_id:MSJ3.8"

CDS
/codon_start=1
/evidence-not_experimental
39943..40386
/note="gene_id:MSJ3.9
unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01422.1"
/db_xref="GI:11994247"
/translation="MAEFHLPLNTEPDQTEDEFEADTLLIDELRYLALDSR
SNQVATQSDVEIICVLICNSSSCSRWLNAGEHLPVVEFPAHEMFKGLVCAICRE
ELAAHRLISLPCKRHHYHKQISNMLSNNTQCHIRHAYELPNHC"
42049..47107
/note="gb|AAD15542.1
gene_id:MSJ3.10
strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
complement(join(47610..47819,47908..48078,48155..48202))
/note="gb|AAD24021.1
gene_id:MSJ3.11
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01423.1"
/db_xref="GI:11994248"
/translation="MAORSLFQESSDOEELDRMEIRMTRYGERRRDRDQVPIEC
DNKVVATSLDLPVTEKLPFSCPYEISGPOCCGPKRWMTVAICHEFDMIKETSE
MKDILHANKVYESQAKIPLMEKKPTELLKKYHSLNKY"
join(49962..50126,50196..50501)
/note="gb|AAC69115.1
gene_id:MSJ3.12
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01424.1"
/db_xref="GI:11994249"
/translation="MSSNSSENSASTQSVSRNGRLPKSGHGLDVTYTSYAKRPG
REPEKPTRKDDILFKWEDGTEEVAVDSWKLVSVDLSAASVAALELALQAV
KHLKQKVKHSRPNQIMKIMIKICPCIAFTIVIVGILMFKINIPORLILGY"
complement(join(51037..51156,51256..51486))
/note="contains similarity to Mu/Spm-like transposon
protein
gene_id:MSJ3.13"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01425.1"
/db_xref="GI:11994250"
/translation="MVFHIGDGLKVEVYLADRDLPVNLGISTWADKTYTF
KNNQALAKRRKRY"
join(52402..52436,52529..52793,52867..52969,53069..53312,
53412..53436,53529..53793,53867..53969,54069..54312)
/note="gene_id:MSJ3.14"
/codon_start=1
/evidence-not_experimental
/product="replication protein A1-like"
/protein_id="BAB01426.1"
/db_xref="GI:11994251"
/translation="MCCYDGLRDLNCCITCMTICVAVFRMYISPTDENSEFLWVEVE
WGQIQALIESRPFQRYDMLENNQKELIRFGVLENGCFVRSYTRKYMILMNTVE
TSSDREVLNRPEDYIIEPTATLTELVDGALVDGCLRNITEDGWMDEGRL
TEIMDKYKEKRLKCEAGELAVDHRKRSFNSISYINDIFVALIMMKLFFYHGLK
HKRPQVAGISDVTVPDNIQVEQELKNYE"

BASE COUNT 18008 a 10125 c 10679 g 16927 t
ORIGIN

Query Match 93.3% Score 14: DB 13: Length 55739;
Best Local Similarity 100.0%; Pired. No. 2.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 qc9qqtatallaga 15
 |||||
 Db 54327 CCGCCTCTATTACA 54340

RESULT 10
 AB024037 81414 bp DNA PLN 09-AUG-2000
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone:MSK10.
 DEFINITION
 AB024037
 VERSION
 AB024037.1 GI:4519196
 KEYWORDS
 Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
 clone:MSK10.
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 11;
 Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 AUTHORS
 TITLE
 Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H.
 and Tabata,S.
 Structural analysis of Arabidopsis thaliana chromosome 5. X.
 Sequence features of the regions of 3,076,755 bp covered by sixty
 P1 and TAC clones
 JOURNAL
 MEDLINE
 20181125
 REFERENCE
 AUTHORS
 TITLE
 Nakamura,Y.
 Submitted (24-FEB-1999) to the DDBJ/EMBL/GenBank databases.
 Yasuoka Nakamura, Kazusa DNA Research Institute, Department of
 Plant Gene Research, 1532-3, Yada, Kasarazu, Chiba 292-0812, Japan
 E-mail:yknakmu@kazusa.or.jp, Tel:81-438-52-3935,
 Fax:81-438-52-3934
 Address for correspondence: kasu@kazusa.or.jp
 For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/gqd.graph.cgi?c=MSK10
 Genes with similarity to proteins in the databases are described in
 product or note qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Graal
 (Informatics Group, Oak Ridge National Laboratory,
 http://complib.ornl.gov/Graal-1.3/),
 GENSCAN (Chris Burge, MIT, http://cchr-081.mit.edu/GENSCAN.html),
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://gremli1.zool.lasalle.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
 http://genome.wustl.edu/coddy/tRNAscan-SE/).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is MG623 and the 3' clone is T13612.

FEATURES
 source
 1.81414
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MSK10"
 /clone_lib="Mitsui P1"
 complement(1..192)
 /note="CDS is reported in Acc# AB028613
 gene_id:MG623.7"
 /number=2
 /evidence="not_experimental"
 /product="retroelement pol polypeptide-like"
 complement(241..2797)
 /note="CDS is reported in Acc# AB028613
 gene_id:MG623.7"
 /number=1
 /evidence="not_experimental"

exon
 exon

CDS
 /product="retroelement pol polypeptide-like"
 /note="emb|CA877928.1
 gene_id:MSK10.2
 similar to unknown protein"
 /codon_start=1
 /evidence="not_experimental"
 /protein_id="BA97394.1"
 /db_xref="GI:8843868"
 /translation="MIRIALHAYCISLPSGLFLPASKLEFTSONKSTPTSDP
 CALIPHTSTSSNLSGHSMTLLDPLVSHATPDPTTPPTLPRTSTWSSITTH
 HHHSSTGHSITSSOSTQIRTPDKLSTRKRGSDKVGSGDLHRTKPLSAHYHSG
 RAIRLLARLSEFYMYR"
 join(4695..5288,5361..5486,5645..5724,5835..5960,
 6033..6137,6227..6328,6399..6458,6542..6629,6876..7058)
 /note="emb|CA81580.1
 gene_id:MSK10.3
 strong similarity to unknown protein"
 /codon_start=1
 /pseudo
 /evidence="not_experimental"
 7427..8119
 /note="contains similarity to mutator-like transposase
 gene_id:MSK10.4"
 /codon_start=1
 /pseudo
 /evidence="not_experimental"
 8890..9582
 /note="contains similarity to mutator-like transposase
 gene_id:MSK10.5"
 /codon_start=1
 /evidence="not_experimental"
 /protein_id="BA97395.1"
 /db_xref="GI:8843869"
 /translation="MPNPIHPDILARIKKMAKCMYLCMKSGRCHDLYPRD
 VLDATIRKCDPDPDYMAIDPDNDSDHRRPKKQDAGNRTTHIEGLRVT
 HESDIGAILHLERLAPRNAAFTCALILICAGNMGVYLRUGSNHVALESEA
 KDICEVEDIKRYGNILKYTANGFEPGCGVSTPKCAEMCYRSRLTNLCOCY
 IMICAKRISQIL"
 join(10112..10220,10432..10664)
 /note="gene_id:MSK10.6
 similar to unknown protein"
 p11102599
 /codon_start=1
 /evidence="not_experimental"
 /protein_id="BA97396.1"
 /db_xref="GI:8843870"
 /translation="MKPGSTYMOYKQTSQSPVKRREYITDVGAIMRGIDLPPI
 SLVNTLPVKQTPLETLRLNLYSSDINTLAIDEMTWSLLOPESTRPYFELLVERG
 SLKVLAKKATSS"
 complement(11072..11536)
 /note="gb|AA74957.1
 gene_id:MSK10.7
 strong similarity to unknown protein"
 /codon_start=1
 /pseudo
 /evidence="not_experimental"
 join(17864..18351,18409..18718,18998..19184,19258..19718
 20145..20803,20889..20997,21270..21941)
 /note="gene_id:MSK10.8"
 /codon_start=1
 /evidence="not_experimental"
 /product="retroelement pol polypeptide-like"
 /protein_id="BA97397.1"
 /db_xref="GI:8843871"
 /translation="MSEDAKREYPPRIYEGDSNLEGGINHNITLCEPFAIESIGK
 ELVAELKTELRLAKLVDYSFLWSKTYVYLLCQRLKLEKMLCTVAAPKPIRTGLN
 EFHITGLINIDELPEKFEVPAADYKAFSEHLLVGEENIDELRQGLVTCRAMPEK
 RRMVIPESAKRYVDKAMKTYPMGARAAYFALVYSIKLRPLIGKITYSGLVFILOMA
 YSELITIAEFGNAVADEILILRWGNGTRATITSVIAEDIKANGELRPTWPKKA
 EISELPHWPDLEDPALDNLITPLHGYKLVINGYDYVTSKKKKLLKDSRRKKKK
 KKKKKYAVSEDEDEKRSRQDGGDGSGLDVVSNLITSPNREFVDTSTKSSS
 RLVTICQLESSVEAKFEGRGSLIENDVKOIKELKAIADSSSYTRDMLAKPOPO

CDS
 /note="gene_id:MSK10.9"
 /codon_start=1
 /evidence-not_experimental
 /protein_id="BA97398.1"
 /db_xref="GI:8843872"
 /translation="MAAQRSLFNGESSVDEELNMEITRMETRYDKRRRDRKGVPIED
 CNAKVVATSLDPTVTKLEFSSCPHEISDGPQGGKGRWMTALDDDEDMKEETTE
 MKDLKANKVESQTEKIFLMEKKEETLEKYSLSKRL"

CDS
 /note="gene_id:MSK10.10"
 /codon_start=1
 /evidence-not_experimental
 /protein_id="BA97399.1"
 /db_xref="GI:8843873"
 /translation="MIMIKAROKOOSLILNPHWPSILITSSPSYTIACNASSS
 VETSDIALANSLSKPANPSOLSTWTSYSSSLHKNKSCIPNNHKN"
 complement(join(25379..25591,25667..25897))

CDS
 /note="gene_id:MSK10.11"
 /codon_start=1
 /evidence-not_experimental
 /product="pn/spm-like transposon protein-like"
 /protein_id="BA97400.1"
 /db_xref="GI:8843874"
 /translation="VPHIDGQFLKQENAVLTADRGIGTWDKTTVTFKRNHOLAV
 AIGDILQNFHAWPTIRMPVYFKAMFELTRHMDSDHVTWTOFNLVARTLY
 RDNMVLKRGWARGIDKPRMTTVMNDIYMPTEFDDPYGMS"
 join(26245..26379,26472..26736,26817..26919,27018..27261)

CDS
 /note="gene_id:MSK10.12"
 /codon_start=1
 /evidence-not_experimental
 /product="replication protein A1-like"
 /protein_id="BA97401.1"
 /db_xref="GI:8843875"
 /translation="MGYDCLADI NPY IICWFLCWVLRKYVSF IITNSYKLVAVVD:
 NGIOTFATIGRSQRYNNLEHNOKEIRTFYIENKGCVRITTKTETERNATEL
 NISDPRENALNRTPTDTIIEHTAHENIYDVALVEWGLRNTPEWMEKMLREL
 TRESMDKREKRGELGELAVDILKKEKSPNDNIYLDNIYALTMKLFVYHLPN
 HAKIQNHAKI SIVATVNNIHEVDELKRVLE"
 complement(join(27691..28087,28296..29179,29384..30543))

CDS
 /note="gene_id:MSK10.13"
 /codon_start=1
 /evidence-not_experimental
 /product="maltotriose-like transposase"
 /protein_id="BA97402.1"
 /db_xref="GI:8843876"
 /translation="MAGPAPVIVISLWKKQKRVFNPDEKGRVQVLPAPPAVYDLSI
 LKAVYDYGVSSTHDLPLSYMSKILKTPEDIPPYISNDROLQYGLTKREOL
 RLVEFTAKDSISIRKDSRGKSRALADSRASREVLWGEVGYCYQYDDEA

Query Match 93.3%; Score 14; DB 12; Length 81414;
 Best Local Similarity 100.0%; Prcd. No. 2,4e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 gcgggtctataga 15
 |||
 Db 28146 GCGGCTCTATTACA 28159

RESULT 11
 AF058825/c

LOCUS AF058825 110157 bp DNA PLN 15-Apr-1998
 DEFINITION Arabidopsis thaliana BAC F7N22.
 ACCESSION AF058825
 VERSION AF058825.1 GI:3047060
 KEYWORDS
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:
 Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids 1:
 Brassicales: Brassicaceae: Arabidopsis.
 1 (bases 1 to 110157)
 Washington University Genome Sequencing Center.
 The A. thaliana genome sequencing project.
 Unpublished (1997)
 2 (bases 1 to 110157)
 Dente, M.
 The sequence of A. thaliana F7N22
 Unpublished (1998)
 3 (bases 1 to 110157)
 Waterston, R.
 Direct Submission
 Submitted (09-APR-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 Submitted by:
 Genome Sequencing Center,
 Department of Genetics, Washington University,
 St. Louis, MO 63108, USA
 e-mail: twatson@wustl.wustl.edu

COMMENT

MAPPING: Clones were assigned to the YAC map by hybridization by
 M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
 by M. Matra, WashU, to pick the best candidates for sequencing.

NOTE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The actual start of this clone is at base position 1 of F7N22;
 actual end is at 110157 of F7N22.

NOTES:

FEATURES
 source
 1..110157
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="IV"
 /clone="F7N22"
 /map="unknown"
 3032..3821
 /note="Pseudogene; similar to reverse transcriptases"
 complement(6554..8653)
 /gene="F7N22.18"
 complement(join(6554..6961,7363..7584,8127..8253,
 8487..8653))
 /gene="F7N22.18"
 /codon_start=1

misc-feature
 gene
 CDS

```

/evidence-not_experimental
/protein_id-"AAC13576.1"
/db_xref-"GI:3047062"
/translation-"MGLLMSKINRKASDKANPISRRNNLIDPHEIDSPGHASTI
RIDPMPRPOTLWFNESSVGVVCSIFERKDFDPHAWTNPQVITCMETPACKT
EETTPDLVLDETHRADRTIDGKSDIYKEVSSRIAEETOLCSGESSEISAGG
LSVQMMNITVHAPRRKRIYVGSILHESASHTCLPEPRDPVLSKLAIVENC
IANQAKINSFGEYEDLADKDFPAIRFGASSTRTPTSANRPREVAANLREVAANL
KMAKHQILLMALAIDKIPVLDLHQHNSN"
11245..14802
/misc_feature
/feature="pseudogene; similar to reverse transcriptases"
/feature="pseudogene; similar to reverse transcriptases"
27112..28030
/gene-"F7N22.3"
/join(27112..27748,27822..28030)
/gene-"F7N22.3"
/feature="contains similarity to ppytidi-prolyl cis-trans
isomerase (Pfam: pro_isomerase; Jmm. Score: 23.86 and
28.41; coded for by A. thaliana cDNA N97251; coded for by
A. thaliana cDNA T43563"
/codon_start=1
/protein_id-"AAC13578.1"
/db_xref-"GI:3047064"
/translation-"MASSSILIPPLIFRNLLSTTATVTHPPPAKPSPDITIDK
VPLDSELPETFRSDPSATLSTPDSSTPLGRVILGLYGRVPLTVSTFKMKTSS
STSKYKTPVHKIFPGQYFLAGRGGGKRDYAEVSLRDEPRNDVYNSKAFLLIAR
AGVSLTSENDDDDIRLDEYRNEFLITGGPSPQLGSGNIYGVILEGLDVT
SISLPTKPSKNIKQNDPAHFIQHRQAKMSLNNRPKTVFISGCHLKVTPSL
SPILP"
32752..33612
/gene-"F7N22.4"
/join(32752..33937,33015..33103,33358..33424,33502..33612)
/gene-"F7N22.4"
/feature="contains similarity to human OS-9 precursor
(GB:U41635)"
/codon_start=1
/evidence-not_experimental
/protein_id-"AAC13579.1"
/db_xref-"GI:3047065"
/translation-"MVMIDKHGNFLCYLPKEKAKSGTSSQONISYVMMETQOIVK
LKPDELQPLSEKLFPROGMSYEHQKRYKQIHVHDDNKQPYWASHVYVNOT
TCDLQSPKRVVPRVCAFRAMVISTHLSICKKALIVGCTTICKH"
complement(35192..33289)
/feature="pseudogene; similar to reverse transcriptases"
complement(39286..40766)
/gene-"F7N22.17"
/feature="pseudogene; similar to reverse transcriptases"
complement(39286..40383,40614..40766)
/gene-"F7N22.17"
/feature="similar to Arabidopsis thaliana re1roltransposon
Tall-1 (GB:U47193)"
/codon_start=1
/evidence-not_experimental
/protein_id-"AAC13575.1"
/db_xref-"GI:3047061"
/translation-"MAMSEDFLGHVPLVHQLISPRMSAFPTIGIKETLMMKGY
SLPNSNAYVRRSVYVTTYNPRKONRLALIGQSRWGGPPDSYVRIEKRVPK
FQTEAMNGLVLRGFWSPNDMMSLHRRVPIHSELEKIPVVOITGIPLEPTAM
ARCVNRLGIVSVDPDENSNHGVVVKIDMMIDLPKHPNQAFADNHYIKVIR
ERLNPSCOSGLKSHYKECVLADPDPAKISDDNDNDIDNOCNKAKKPSIDPITQ
TVPDPAKIPGLPQPSYCNKQNTONTVSVSPSEVPTDITVRIKYLAKIARVET
OPFISNDLLESDSSNAONQFVLLKRRKVEEDMYKRVKVAEDMAMVLSQLKKKKKES
AGSCSTHNGLDGAGGPPRPDP"
46610..47864
/misc_feature
/feature="pseudogene; similar to retrovirus-related pol,
polyproteins"
polyproteins
/gene-"F7N22.6"
/join(57941..58149,58209..59137,59251..59852)
/gene-"F7N22.6"
/feature="contains similarity to retrovirus-related pol,
polyproteins"
/codon_start=1
/evidence-not_experimental
/protein_id-"AAC13580.1"

```

```

/db_xref-"GI:3047066"
/translation-"MTDINTNSKIVGNTVETPDEVALHLAKELEKTOITDSQMDV
RRRI.PANDSTNTTIPSSAQCMPPDINMRVYTEDDPPIVPGCTTYVTVGCMITMAQ
GIVYCKOLLINLVNULFDLYNKAQSSKTLILKRYKTDSCGQORSTAKPLNTKM
VDSKPIMEVEVALQHSIPILBQMSICWFKNCILIKLIPQCHIDKNTLNKRRAM
TFPDLVRKPIEGNNYGADAGQNDHVNADHAKLKGKRGVSTIPQKLVSSIT
NFKRSPERKFGKCHGCKLGHKGDYCKKADYKSAQNTLEEDMAVYTECMDD
NOVENYEDGCTTHICTDRTMETSTYVKNSEDLFGKNTAMSMIEGGLVKSIVTEP
KNVASVYENKNNPITVIVSPITVMEHRIQVNYKIMRQNMNI.IPKFTNDEKCE
VCYHAKITKPSPRVETTEPMNLIPKFTNDEKCEVCYHAKITKPSPRVETTEP
GLVITDLCIDIKVQTRKCKRPVPIIDCTRTTYVYLLMSKQBALVKKRHTLVENO
LQRTIKVSDRG"
63170..64252
/gene-"F7N22.7"
/feature="pseudogene; similar to Arabidopsis thaliana re1roltransposon
Alula (GB:X81801)"
/gene-"F7N22.7"
/feature="similar to Arabidopsis thaliana re1roltransposon
Alula (GB:X81801)"
/codon_start=1
/evidence-not_experimental
/protein_id-"AAC13581.1"
/db_xref-"GI:3047067"
/translation-"MPLNTVYKVIHQDSASISAPVVTGLPKSIQNKVYATAHAT
ICHDRELTPHVEDLITGNDVQDEASTIYEVSEENISQFHLIGSTSEKAT
IKRIVKFKETPLPSALPWFERRAMEKRSATKQDEIVAMPLEVLYLIHDPH
KVDNLTLEKINNYQULDESDAPPSAUKRIVQKRLDPSGFTLLPSIGELAFSDC
LQVIGAVYSIMPLISVARRLEIQRKPCILILADRSYKPKPGMLKQIPMNGVEVP
TNYVLDMEYEHKDPILICRPHIASGVGVVAVHCKISLIDKHNKIQPDINKIPQNS
TVDKTSKQRYVHCGCTRTYKAKKA"
64919..67482
/gene-"F7N22.8"
/join(64919..64986,66088..66270,66336..66473,66723..67482)
/gene-"F7N22.8"
/codon_start=1
/evidence-not_experimental
/protein_id-"AAC13577.1"
/db_xref-"GI:3047063"
/translation-"MIONKETDEGFTAAAMSPTLPDEFERRAARRQRAEJARCK
KMSRYELIDEDIKTEYEPHSMKFKILINKSDVYVPEIPVQVGGAAVRSYSS
KETSIFELSLQVIMYHGLIPELITMCHSGIVSNIDKMLISALILILITGKGV
IKQILNRPVPMVLIHICVYKMAKLNCKKRYKALCVGVVPLVAVCVPLKKEV
LAPAMLDILRCESEEDRYGUPHIREHHSIRANILPFCIATIRILEGNIDF
KPADELIFEGSPPEISHEGATIDEDDTVIDAEFDSHFSHPKPAKSK
SLSAIDNNNSKRLQKCKODKLLAKCLAKIKFLDKISCSSTTTPQ"
72089..73326
/misc_feature
/feature="pseudogene; similar to retrovirus-related pol,
polyproteins"
complement(84219..86759)
/gene-"F7N22.15"
complement(84219..86759)
/gene-"F7N22.15"
/feature="contains similarity to retrotransposon-like
proteins"

```

Query Match 93.3%; Score 14; DB 12; Length 110157;
 Best local similarity 100.0%; Pred. No. 2,4c102;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 gccgcgtctatlaaa 15
Db 92562 gccgcgtctatlaaa 92549

RESULT 12
LOCUS F10A2 123080 bp DNA PLN 13-ATG-1999
DEFINITION Arabidopsis thaliana BAC F10A2.
ACCESSION AF147259
VERSION AF147259.1 GI:4732164
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana

```


misc_feature /note="assembly_fragment:01928
fragment_chain:2"
78437..84166
/note="assembly_fragment:01930
fragment_chain:3"
84267..96814
/note="assembly_fragment:01918
fragment_chain:3"
96915..118894
/note="assembly_fragment:01110
fragment_chain:3"
118995..121637
/note="assembly_fragment:01381
fragment_chain:4"
121738..129786
/note="assembly_fragment:01421
fragment_chain:4"
129887..133215
/note="assembly_fragment:00105"
133316..138221
/note="assembly_fragment:00370"
138322..154648
/note="assembly_fragment:01029"
154749..156946
/note="assembly_fragment:01170"
159047..166727
/note="assembly_fragment:00541
fragment_chain:5"
166828..173339
/note="assembly_fragment:00224
fragment_chain:5"
173440..175742
/note="assembly_fragment:00840
fragment_chain:5"
clone_end:SP6
vector_side:right"
BASE COUNT 52688 a 33525 c 33326 g 54000 t 2203 others
ORIGIN

Query Match 93.3%; Score 14; DB 81; Length 175742;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 qcqggtctatlaqa 15
|||||
Db 31936 GCGGCTATTAGA 31923

Search completed: May 1, 2001, 17:45:29
Job time: 6945 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 16:41:49 ; Search time 113.81 Seconds

(without alignments)
76.942 Million cell updates/sec

Title: US-09-593-914-1

Sequence: 15
1 agcgcgcctatata 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_0401:*

1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:*

2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:*

3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:*

4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:*

5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:*

6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:*

7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:*

8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:*

9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:*

10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:*

11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:*

12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:*

13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:*

14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:*

15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:*

16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:*

17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:*

18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:*

19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:*

20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:*

21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:*

22: /SID56/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13.4	89.3	1071	21	Arabidopsis thaliana
2	13.4	89.3	1085	21	Arabidopsis thaliana
3	13.4	89.3	3408	19	Dead Box X (DBX) g
4	13.4	89.3	5322	21	Human DBX1 coding
5	13.4	89.3	5322	19	Dead Box X (DBX) g
6	13.4	89.3	8957	16	HTLV-II virus NRA
7	13.4	86.7	360	11	Sequence encoding
8	13.4	86.7	396	11	Sequence encoding
9	13.4	86.7	843	11	Sequence encoding
10	13.4	86.7	1151	21	Aspergillus oryzae
11	13.4	86.7	1260	13	Sequence encoding

12	13	86.7	1782	17	T38241
13	12.4	82.7	533	20	X21072
14	12.4	82.7	1167	18	T79770
15	12.4	82.7	1167	18	T79791
16	12.4	82.7	1336	21	C69404
17	12.4	82.7	1407	21	C65280
18	12.4	82.7	1559	21	Z47168
19	12.4	82.7	1815	21	Z47167
20	12.4	82.7	2022	17	T16016
21	12.4	82.7	2127	17	T12414
22	12.4	82.7	2249	13	Q29710
23	12.4	82.7	2832	15	O62186
24	12.4	82.7	4809	15	O62185
25	12.4	82.7	7475	12	Q15312
26	12.4	82.7	11475	19	V54979
27	12.4	82.7	11478	18	T79784
28	12.4	82.7	11478	18	T79805
29	12.4	82.7	11478	18	V55069
30	12.4	82.7	13811	20	X20649
31	12.4	82.7	143068	21	F21105
32	12.4	82.7	143068	21	F21272
33	12.4	82.7	143068	21	A34983
34	12.4	82.7	143068	21	A35150
35	12.4	82.7	149412	21	A35151
36	12.4	82.7	152740	21	F21273
37	12.4	82.7	1830121	17	T42063
38	12	80.0	296	21	A79502
39	12	80.0	970	21	C47429
40	12	80.0	973	21	C33442
41	12	80.0	1167	19	Z96423
42	12	80.0	1167	19	V37401
43	12	80.0	1486	21	C51465
44	12	80.0	1488	21	C33887
45	12	80.0	1491	21	C76525

ALIGNMENTS

RESULT 1	
C46237	
ID	C46237 standard; DNA; 1071 BP.
AC	C46237;
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 49404.
XX	
KM	Hybridisation assay; genetic mapping; gene expression control;
KW	protein identification; signal transduction pathway;
KM	metabolic pathway; promoter; termination sequence; ss.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
XX	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.

Mycoplasma 46-48 k
Polynucleotide seq
Potato starch bran
Potato starch bran
Human secreted pro
Human secreted pro
Mouse sphingosine
Mouse sphingosine
DNA encoding ubiq
Ub17 ubiquitin-1yt
PGK promoter from
Cyanobacteria repl
pAO1 plasmid fragm
Plasmid pHP12 enc
Potato starch bran
Full length potat
Potato class B sta
Polynucleotide seq
Human low adenosi
Human low adenosi
Human adenosine re
Human adenosine re
Human adenosine re
Human low adenosi
Haemophilus influe
Pinus radiata cell
Arabidopsis thalia
Arabidopsis thalia
S. pneumoniae deri
Streptococcus pneu
Arabidopsis thalia
Arabidopsis thalia
Human ORF2080

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131444.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132481.
PR 05-MAY-1999; 99US-0132481.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132861.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134766.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138447.
PR 14-JUN-1999; 99US-0139111.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139450.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142830.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152463.
PR 10-SEP-1999; 99US-0153707.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 89.3%; Score 13.4; DB 21; Length 1071;
Best Local Similarity 93.3%; Pred. No. 41;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcgggtctattaga 15
|||
Db 922 agtgggtctattaga 936

RESULT 2
ID C33698 standard; DNA; 1085 BP.
XX
AC C33698;

DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 3998.

XX
KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.

XX
OS Arabidopsis thaliana.
XX
PN EP103405-A2.
XX
PD 06-SEP-2000.

XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0133149.
PR 30-APR-1999; 99US-01332048.
PR 30-APR-1999; 99US-01332047.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

```

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147303.
PR 04-AUG-1999; 99US-0147308.
PR 04-AUG-1999; 99US-0147304.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147392.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148371.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148345.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151060.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

```

```

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 89.3%; Score 13.4; DB 21; Length 1085;
Best Local Similarity 93.3%; Pred. No. 41;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 agcgggtctataga 15
Db 919 agtgggtctataga 933

```

```

RESULT 3
ID V69632 standard; DNA: 3408 BP.
V69632;

```

```

AC V69632;
XX 02-FEB-1999 (first entry)
DT
XX

```

```

DE Dead Box X (DBX) gene short transcript nucleic acid sequence.

```

```

XX Non-recombining region; human: Y chromosome; X homologue; testis; DBX;
XX infertility; sperm; gene alteration; inhibitor; Dead Box X; ss.
XX

```

```

OS Homo sapiens.

```

```

PN W09846747-A2.

```

```

PD 22-OCT-1998.

```

```

PF 10-APR-1998; 98WD-US07115.

```

```

PR 11-APR-1997; 97US-0041877.

```

```

PA (WHD) WHITEHEAD INST BIOMEDICAL RES.

```

```

PI Lahn BT, Page DC;

```

```

DR WPI: 1998-568729/48.

```

```

PT P-FSDB; W81502.

```

```

PT Novel genes in the non-combining region of Y chromosome - useful to

```

```

PT diagnose if male infertility or reduced sperm count has a genetic

```

```

PS basis

```

```

PS Disclosure: Fig 3A-B; 54pp; English.

```

```

XX This represents the nucleotide sequence of the Dead Box X (DBX) gene

```

```

XX short transcript. The invention relates to genes occurring on the non-

```

```

XX recombining region of the human Y chromosome. The sequences fall into two

```

```

XX classes: (1) X-homologous DNA which are expressed in many organs, having

```

```

XX functional X homologies and (2) testis-specific DNA sequences. Y

```

```

XX reduced sperm count can be assessed using the invention to determine

```

```

XX whether the condition is associated with or caused by the occurrence of

```

```

XX the gene or gene alteration. Candidate inhibitors of the enzymatic

```

```

XX activity of the genes can be assessed using in vitro assays.

```

```

XX Sequence 3408 BP; 951 A; 734 C; 923 G; 800 T; 0 other;

```

Query Match 89.3%; Score 13.4; DB 19; Length 3408;
Best Local Similarity 93.3%; Pred. No. 47;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 agcgggtctattaga 15
||||| |||||
Db 728 agcgggtctattaga 742

RESULT 4
A29207
ID A29207 standard; DNA; 5321 BP.
XX
AC A29207;
XX
DT 12-SEP-2000 (first entry)
XX
DE Human DBX1 coding sequence.
XX
KW Topoisomerase III-alpha; interacting protein; TIII-a IP; RNA helicase;
KW DBX1; nucleosome; supercoiling; chromosome segregation; recombination;
KW stabilization; cell division; apoptosis; cell cycle regulation;
KW cytotoxic; anti-tumour; DBX1; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 856..2844
FT /*tag= a
FT /standard_name= DBX1
FT /product= RNA_helicase-like_protein

XX PN M0200032768-A1.
XX 08-JUN-2000.
XX PD 29-NOV-1999; 99WO-FR02952.
XX PF 30-NOV-1998; 98FR-0015081.
XX PR (AVET) AVENTIS PHARMA SA.
XX PA Pournier A, Goulaouic H, Riou J;
XX P1 WPI: 2000-412316/35.
XX DR P-PSDB; Y96483.
XX XX
XX PT New nucleic acid encoding ligand for topoisomerase III alpha, useful
XX for inhibiting the enzyme and in drug screening, e.g. for potential
XX anticancer agents
XX PS
XX PS Claim 6: Page 47-53; 68pp; French.

CC This DNA encodes a DBX1, which shows homology with RNA helicases but the
CC activity of a helicase has never been demonstrated and its function has
CC not yet been identified. DBX1 possesses the 8 characteristic motifs of
CC the helicases of the family "DEAD". In particular, it appears to be part
CC of the sub-family represented by helicase PL10. The DBX1 gene is situated
CC on the X chromosome and its homologue, which is situated on the Y
CC chromosome has 91 percent identity with a novel protein coding sequence.
CC The novel protein (see Y96482) is a human topoisomerase III-alpha
CC interacting protein (TIII-a IP). TIII-a IP has structural features in
CC common with RNA helicases (e.g. DBX1), which are involved in destruction
CC of the nucleosome structure, supercoiling of DNA, segregation of newly
CC replicated chromosomes, and recombination and stabilization of the
CC genome. Agents that alter interaction between TIII-a IP and TIII-a
CC therefore modulate cell division, replication, transcription,
CC translation, splicing and DNA recombination or repair, so may slow cell
CC growth, block the cell cycle or induce apoptosis. Antibodies and ligands
CC of TIII-a IP are used to prevent, treat or alleviate diseases that
CC involve abnormal regulation of the cell cycle, i.e. they are potential

CC anti-tumour (cytotoxic) agents.

XX SQ Sequence 5321 BP; 1532 A; 1010 C; 1274 G; 1505 T; 0 other;

Query Match 89.3%; Score 13.4; DB 21; Length 5321;
Best Local Similarity 93.3%; Pred. No. 49;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 agcgggtctattaga 15
||||| |||||
Db 727 agcgggtctattaga 741

RESULT 5
V69631
ID V69631 standard; DNA; 5322 BP.
XX
AC V69631;
XX
DT 02-FEB-1999 (first entry)
XX
DE Dead Box X (DBX) gene long transcript nucleic acid sequence.

XX KW Non-recombining region; human; Y chromosome; X homologue; testis; DBX;
KW infertility; sperm; gene alteration; inhibitor; Dead Box X; ss.
XX
OS Homo sapiens.
XX
PN M09846747-A2.
XX PD 22-OCT-1998.
XX
PP 10-APR-1998; 98WO-US07115.
XX
PR 11-APR-1997; 97US-0041877.
XX
PA (WHEE) WHITEHEAD INST BIOMEDICAL RES.

XX P1 Iahn BT, Page DC;
XX DR WPI: 1998-568729/48.
XX DR P-PSDB; W81501.

XX PT Novel genes in the non-combining region of Y chromosome - useful to
XX diagnose if male infertility or reduced sperm count has a genetic
XX basis
XX PS
XX PS Disclosure: Fig 3A-B; 54pp; English.

CC This represents the nucleotide sequence of the Dead Box X (DBX) gene long
CC transcript. The invention relates to genes occurring on the non-
CC recombining region of the human Y chromosome. The sequences fall into two
CC classes: (1) X-homologous DNA which are expressed in many organs, having
CC functional X homologues and (2) testis-specific DNA sequences. Y
CC chromosomal DNA from males with known conditions such as infertility and
CC reduced sperm count can be assessed using the invention to determine
CC whether the condition is associated with or caused by the occurrence of
CC the gene or gene alteration. Candidate inhibitors of the enzymatic
CC activity of the genes can be assessed using in vitro assays.

XX SQ Sequence 5322 BP; 1532 A; 1011 C; 1274 G; 1505 T; 0 other;

Query Match 89.3%; Score 13.4; DB 19; Length 5322;
Best Local Similarity 93.3%; Pred. No. 49;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 agcgggtctattaga 15
||||| |||||
Db 728 agcgggtctattaga 742

PA	(ABBO) ABBOTT LAB.
XX	
PI	Buytenroep NH, Chan EM, Chen ISY, Edwards M, Golde DF;
PI	Guldinger F, Idler KB, Johnson JE, Lee HH, Molley CT;
P1	Peterson B, Robertson E, Rosenblatt JD, Stephens JE;
P1	Swanson PA, Tate C;
XX	
DR	WPI; 1995-061020/08.
DR	P-PsDB; R64971/77.
XX	
P1	Human T-cell lymphotropic type II NRA viral genome, - expression
P1	products, and compositions for use in improved methods for detection
PT	of HIV-1 infection
PS	Claim 1; Page 39-45; 10Opp; English.
XX	
CC	The complete sequence of the HTLV-II provirus NRA-19a strain genome is
CC	presented. The gag region encodes proteins p19, p24 and p15, and
CC	the pol region (including the pol precursor) and env region
CC	(encoding gp46 and p21e) are indicated. The tax/rex region comprises
CC	the tax region (encoding p40x) and the rex region (encoding p26),
CC	which partially overlap, and are involved in regulating or controlling
CC	gene expression. The virus may be produced in cell culture ATCC
CC	CRI, 11560, and the encoded proteins may be produced as recombinant
CC	fusion proteins. The DNA and gene products may be used in
CC	diagnostic methods for detection of HTLV-II DNA or antibodies in test
CC	samples, by immunosay, liase chain reaction, polymerase chain
CC	reaction, Southern blot hybridisation or Western blot methods, and in
CC	antibody and vaccine preparation.
SQ	
XX	Sequence 8957 BP; 2154 A; 3208 C; 1648 G; 1947 T; 0 other;
QY	
Db	Best Local Similarity 89.3%; Score 13.4; DB 16; Length 8957; Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0 1 agcggagctattaga 15 6653 AGCGGCTCTATTATA 6639
RESULT	7
ID	Q05577 standard; DNA; 360 BP.
AC	Q05577:
XX	
DT	19-DEC 1990 (first entry)
XX	
DE	Sequence encoding swine enzootic pneumonia mycoplasma surface
DE	antigen polypeptide.
XX	
KW	Mycoplasma pneumoniae; enzootic pneumoniae ds;
OS	Mycoplasma hyopneumoniae.
PN	JB02167079-A.
PD	27-JUN-1990.
PF	21-DEC-1988; 88JP-0322829.
PR	21-DEC-1988; 88JP-0322829.
PA	(NISE-) NIPPON SEIPUN KK.
XX	
WI	WPI; 1990-241949/32.
DR	P-PsDB; R06278.
XX	
PT	Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
PT	for early detection and treatment of mycoplasma or enzootic
TT	pneumonia of pigs

XX Claim 1; Page 570; 28pp; Japanese.
PS
XX Sequence encoding polypeptides may be used as probes for the early
CC diagnosis of pneumonia in pigs. The sequences may be used to produce
CC an expression vector and transform a suitable host.
XX
XX
SQ Sequence 360 BP; 130 A; 52 C; 64 G; 114 T; 0 other;

Query Match 86.7%; Score 13; DB 11; Length 360;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcgggtctattag 14
|||||
Db 61 gcgggtctattag 73

RESULT 8
Q05576
ID Q05576 standard; DNA; 396 BP.
XX
AC Q05576;
XX
DT 19-DEC-1990 (first entry)
XX
DE Sequence encoding swine enzootic pneumonia mycoplasma surface
DE antigen polypeptide.
XX
KW Mycoplasma pneumonia; enzootic pneumonia; ds;
XX
OS Mycoplasma hyopneumoniae.
XX
PN JP02167079-A.
XX
PD 27-JUN-1990.
XX
PF 21-DEC-1988; 88JP-0322829.
XX
PR 21-DEC-1988; 88JP-0322829.
XX
PA (NISR-) NIPPON SEIFUN KK.
XX
DR WPI: 1990-241949/32.
XX
P-PSDB; R06277.

XX Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
PT for early detection and treatment of mycoplasma or enzootic
PT pneumonia of pigs
XX
PS Claim 2; Page 570; 28pp; Japanese.
XX

CC Sequence encoding polypeptides may be used as probes for the early
CC diagnosis of pneumonia in pigs. The sequences may be used to produce
CC an expression vector and transform a suitable host.
XX
SQ Sequence 396 BP; 145 A; 62 C; 70 G; 119 T; 0 other;

Query Match 86.7%; Score 13; DB 11; Length 396;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcgggtctattag 14
|||||
Db 43 gcgggtctattag 55

RESULT 9
Q05578
ID Q05578 standard; DNA; 843 BP.
XX

AC Q05578;
XX
DT 19-DEC-1990 (first entry)
XX
DE Sequence encoding swine enzootic pneumonia mycoplasma surface
DE antigen polypeptide.
XX
KW Mycoplasma pneumonia; enzootic pneumonia; ds;
XX
OS Mycoplasma hyopneumoniae.
XX
PN JP02167079-A.
XX
PD 27-JUN-1990.
XX
PF 21-DEC-1988; 88JP-0322829.
XX
PR 21-DEC-1988; 88JP-0322829.
XX
PA (NISR-) NIPPON SEIFUN KK.
XX
DR WPI: 1990-241949/32.
XX
P-PSDB; R06277.

XX Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
PT for early detection and treatment of mycoplasma or enzootic
PT pneumonia of pigs
XX
PS Claim 2; Page 570; 28pp; Japanese.
XX

CC Sequence encoding polypeptides may be used as probes for the early
CC diagnosis of pneumonia in pigs. The sequences may be used to produce
CC an expression vector and transform a suitable host.
XX
SQ Sequence 843 BP; 324 A; 127 C; 139 G; 253 T; 0 other;

Query Match 86.7%; Score 13; DB 11; Length 843;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcgggtctattag 14
|||||
Db 103 gcgggtctattag 115

RESULT 10
F13364/c
ID F13364 standard; cDNA; 1151 BP.
XX
AC F13364;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae RST SEQ ID NO:5887.
XX

XX Multiple gene expression; filamentous fungal cell; RST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.

```

PR 27-AUG-1990; 90JP-0224945.
PA (NIFL-) NIPPON FLOUR MILLS.
XX
XX Seto Y, Futo S, Mitsuse S, Matsuo K, Tsuna M;
XX
XX WPI: 1992-089874/12.
XX
XX P-PSDB; R21829.
XX
XX DNA and peptide of mycoplasma hyopneumoniae - useful for
XX diagnosis and treatment of swine mycoplasma pneumonia
XX
XX Claim 2; Page 6 and pages 36-37; 45pp; English.
XX
XX The inventors claim DNA encoding a surface antigen and primers used
XX in a method of diagnosing mycoplasma pneumoniae of swine.
XX Mycoplasma cells were collected from the culture of M.ph, lysed in a
XX buffer containing SDS, followed by the extraction of DNA and
XX purification thereof. The DNA of M.ph is cleaved with a restriction
XX enzyme HindIII. The resulting fragment is inserted into a plasmid
XX pUC119 at the site cleaved with HindIII. This plasmid is transformed
XX into E. coli as a host cell. Colony hybridization is performed by the
XX DNA probe method utilizing EcoRI fragments of pKRM1 and pKRM2 to
XX select the bacterial cells containing a plasmid (pKRM126) carrying the
XX 46 kd antigenic gene. E. coli JM 109 which comprises the plasmid
XX pKRM1 or pKRM2 is deposited with FRI under accession No. FERM P-10318
XX or P-10319. The base sequence of the 46 kd antigen gene is given in
XX Q22042.
XX
XX Sequence 1260 BP; 476 A; 204 C; 212 G; 368 T; 0 other;
XX
XX Query Match 86.7%; Score 13; DH 13; length 1260;
XX Best local Similarity 100.0%; Pred. No. 74;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 2 gcgggctcattag 14
XX ||||||||||||
XX 517 gcgggctcattag 529
XX
XX Db
XX
XX RESULT 12
XX T38241
XX ID T38241 standard; DNA; 1782 BP.
XX
XX T38241;
XX AC
XX XX
XX XX 19-DEC-1996 (first entry)
XX DF
XX DE Mycoplasma 46-48 kDa protective antigen gene.
XX XX
XX XX Mycoplasma hyopneumoniae strain Beaufort.
XX OS
XX XX
XX XX Antigen: vaccine; mycoplasma pneumoniae; swine enzootic pneumonia;
XX KM diagnosis; antibody; ss.
XX KW
XX XX
XX XX Mycoplasma hyopneumoniae strain Beaufort.
XX XX
XX XX Key Location/Qualifiers
XX FH 523..1782
XX FT /*tag a
XX FT /transl_except- (730..732, aa:Trp)
XX FT /transl_except- (823..825, aa:Trp)
XX FT /transl_except- (1282..1284, aa:Trp)
XX FT /note "TGA is a stop codon"
XX XX
XX XX MO9628472-A1.
XX
XX 19-SEP-1996.
XX PD
XX XX 15-MAR-1996; 96MO-AU00149.
XX PF
XX XX 16-MAR-1995; 95AU-0001789.
XX PR
XX XX
XX XX (UYME ) UNIV MELBOURNE.

```

XX Doughty SW, Lee R, Walker J;
 XX WPI: 1996-433763/43.
 DR P-PSDB; W01037.
 XX
 PT Putative protective antigens against Mycoplasma - used for the
 PT detection, prevention or treatment of Mycoplasma infections, esp. M.
 PT hyopneumoniae in swine
 XX
 PS Claim 14: Page 28-29; 43pp; English.
 XX
 CC The gene (T38241) coding for a 48 kDa putative protective antigen
 CC (W01037) was obcd. from a Mycoplasma hyopneumoniae genomic library
 CC by screening with a probe generated by PCR amplification (see also
 CC T13814-16). The antigen had originally been isolated from M.
 CC hyopneumoniae cells using antibody probes enriched with Mycoplasma-
 CC specific antibodies. Other protective antigens were also identified
 CC (see also W01024-31). Protective antigens and antibodies can be
 CC used in vaccines for preventing or treating mycoplasma infections,
 CC paritig. M. hyopneumoniae infections in swine. They can also be used
 CC for diagnosis.
 SQ
 SO Sequence 1782 BP; 682 A; 254 C; 259 G; 587 T; 0 other;
 XX
 XX
 Query Match 86.7%; Score 13; DB 17; Length 1782;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2 gcgggtctattag 14
 Db 1039 gcgggtctattag 1051
 XX
 RESULT 13
 X21072
 ID X21072 standard; DNA: 533 BP.
 XX
 AC X21072;
 XX
 DT 05-MAY-1999 (first entry)
 XX
 DE Polynucleotide sequence from the genome of Treponema pallidum.
 XX
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
 KW enzyme production; ds.
 XX
 OS Treponema pallidum.
 XX
 PN W09859034-A2.
 XX
 PD 30-DEC-1998.
 XX
 PF 23-JUN-1998; 98WO-US13041.
 XX
 PR 24-JUN-1997; 97US-0050667.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fraser CM;
 XX
 DR WPI: 1999-081273/07.
 XX
 PT New isolated Treponema pallidum nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of T. pallidum infections, particularly syphilis
 XX
 PS Claim 1; Page 1052; 1150pp; English.
 XX
 CC X20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for T. pallidum

CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 XX
 SQ Sequence 533 BP; 131 A; 111 C; 145 G; 143 T; 3 other;
 XX
 XX
 Query Match 82.7%; Score 12.4; DB 20; Length 533;
 Best Local Similarity 92.9%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 2 gcgggtctattag 15
 Db 1 gcgggtctattag 14
 XX
 RESULT 14
 T79770
 ID T79770 standard; DNA: 1167 BP.
 XX
 AC T79770;
 XX
 DT 10-NOV-1997 (first entry)
 XX
 DE Potato starch branching enzyme gene intron 1.
 XX
 KW Starch branching enzyme; SBE; potato; genetic engineering;
 KW transgenic plant; starch production; modification; ss.
 XX
 OS Solanum tuberosum.
 XX
 PN W09704113-A2.
 XX
 PD 06-FEB-1997.
 XX
 PE 12-JUL-1996; 96WO-EP03053.
 XX
 PR 14-JUL-1995; 95GB-0014437.
 XX
 PA (DANT-) DANISCO AS.
 XX
 PI Poulsen P;
 XX
 DR WPI: 1997-132651/12.
 XX
 PT Affecting enzymatic activity in plant by expressing sense intron
 PT sequence - esp. for inhibiting starch branching enzyme to create
 PT modified starch without post-harvest derivatisation
 XX
 PS Claim 8; Page 30; 70pp; English.
 XX
 CC T79770 represents intron 1 of the potato starch-branching enzyme
 CC (SBE) gene. Sequences antisense to introns of the SBE gene are used
 CC to produce transgenic plants producing modified or new forms of starch
 CC for industrial use. The sequences allow starches to be made without
 CC post-harvest modification (which uses hazardous chemicals and requires
 CC much energy). Using the antisense intron sequences is more reliable,
 CC efficient and selective than known methods for regulating enzymatic
 CC activity because there is almost no homology between introns in
 CC different SBE genes. The SBE gene promoter (see T79783) is also
 CC useful in genetic engineering.
 XX
 SQ Sequence 1167 BP; 324 A; 155 C; 228 G; 458 T; 2 other;
 XX
 XX
 Query Match 82.7%; Score 12.4; DB 18; Length 1167;
 Best Local Similarity 92.9%; Pred. No. 1.8e+02;
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 2 gcgggtctattag 15
 Db 537 gcgggtctattag 550


```

RESULT 15
T79791/c
ID T79791 standard; DNA; 1167 bp.
XX
AC T79791;
XX
DT 11-NOV-1997 (first entry)
XX
DE potato starch branching enzyme gene intron 1 antisense sequence.
XX
KM Starch branching enzyme; SBE; potato; genetic engineering;
XX transgenic plant; starch production; modification; ss.
XX
OS Solanum tuberosum.
XX
PN W09704112-A2.
XX
PD 06-FEB-1997.
XX
PF 12-JUL-1996; 96WO-EP03052.
XX
PR 14-JUL-1995; 95GB-0014435.
XX
PA (DAMI-) DANISCO AS.
XX
PI Poulsen P;
XX
DR WP1; 1997-132650/12.
XX
PT Affecting enzymatic activity in plant by expressing anti-sense
PT intron sequence esp. for inhibiting starch branching enzyme thus
PT creating modified starch without post-harvest derivatisation
XX
PS Claim 8; Page 38; 70pp; English.
XX
CC T79791 is an antisense sequence of intron 1 of the potato starch-
CC branching enzyme (SBE) gene. Sequences antisense to introns of the
CC SBE gene were used to produce transgenic plants producing modified
CC or new forms of starch for industrial use. The sequences allow
CC starches to be made without post-harvest modification (which uses
CC hazardous chemicals and requires much energy). Using the antisense
CC intron sequences is more reliable, efficient and selective than
CC known methods for regulating enzymatic activity because there is
CC almost no homology between introns in different SBE genes. The SBE
CC gene promoter (see T79804) is also useful in genetic engineering.
XX
SQ Sequence 1167 bp; 458 A; 228 C; 155 G; 324 T; 2 other;

Query Match      82.7%; Score 12.4; DB 18; Length 1167;
Best Local Similarity 92.9%; Pred. No. 1.8e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 qcgggtctattaga 15
   | ||||| |||||
Db 631 GAGGCTCTATTAGA 618

```

Search completed: May 1, 2001, 18:20:18
 Job time: 5909 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2001, 15:51:57 ; Search time 65.16 Seconds
(without alignments)
40.194 Million cell updates/sec

Title: US-09-593-914-1

Perfect score: 15

Sequence: 1 agcgggtcattaga 15

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*

2: /cgn2_6/prodata/1/ina/5B.COMB.seq:*

3: /cgn2_6/prodata/1/ina/6A.COMB.seq:*

4: /cgn2_6/prodata/1/ina/6B.COMB.seq:*

5: /cgn2_6/prodata/1/ina/PCITUS.COMB.seq:*

6: /cgn2_6/prodata/1/ina/backfiles1.seq:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	13.4	89.3	3408 4 US-09-058-489-14	Sequence 14, Appl
2	13.4	89.3	5322 4 US-09-058-489-13	Sequence 13, Appl
3	12.4	82.7	1167 4 US-08-981-803-1	Sequence 1, Appl
4	12.4	82.7	1167 4 US-08-981-803-15	Sequence 15, Appl
5	12.4	82.7	2022 2 US-08-505-486-96	Sequence 96, Appl
6	12.4	82.7	2022 3 US-08-801-028-96	Sequence 96, Appl
7	12.4	82.7	2022 3 US-09-340-154-96	Sequence 96, Appl
8	12.4	82.7	2022 5 PCT-US95-09338-96	Sequence 96, Appl
9	12.4	82.7	2022 5 PCT-US95-09339-96	Sequence 96, Appl
10	12.4	82.7	2127 2 US-08-505-486-95	Sequence 95, Appl
11	12.4	82.7	2127 3 US-08-801-028-95	Sequence 95, Appl
12	12.4	82.7	2127 3 US-09-340-154-95	Sequence 95, Appl
13	12.4	82.7	2127 5 PCT-US95-09338-95	Sequence 95, Appl
14	12.4	82.7	2127 5 PCT-US95-09339-95	Sequence 95, Appl
15	12.4	82.7	2250 1 US-08-483-639-1	Sequence 91, Appl
16	12.4	82.7	11478 4 US-08-981-803-29	Sequence 29, Appl
17	11.8	78.7	68 2 US-08-984-578-151	Sequence 151, Appl
18	11.8	78.7	347 1 US-08-149-695-3	Sequence 3, Appl
19	11.8	78.7	347 1 US-08-377-228-3	Sequence 3, Appl
20	11.8	78.7	377 1 US-08-149-695-4	Sequence 4, Appl
21	11.8	78.7	377 1 US-08-377-228-4	Sequence 4, Appl
22	11.8	78.7	464 1 US-08-149-695-5	Sequence 5, Appl
23	11.8	78.7	464 1 US-08-377-228-5	Sequence 5, Appl
24	11.8	78.7	595 1 US-08-149-695-6	Sequence 6, Appl
25	11.8	78.7	595 1 US-08-377-228-6	Sequence 6, Appl
26	11.8	78.7	735 3 US-09-135-782-3	Sequence 3, Appl
27	11.8	78.7	1000 1 US-08-117-083-63	Sequence 63, Appl

28	11.8	78.7	1579 1 US-08-149-695-7	Sequence 7, Appl
29	11.8	78.7	1579 1 US-08-377-228-7	Sequence 7, Appl
30	11.8	78.7	1835 1 US-08-361-611-3	Sequence 3, Appl
31	11.8	78.7	1835 1 US-08-565-655-3	Sequence 3, Appl
32	11.8	78.7	1835 2 US-08-946-967-3	Sequence 3, Appl
33	11.8	78.7	2454 3 US-09-221-235-7	Sequence 7, Appl
34	11.8	78.7	2454 4 US-09-221-235-7	Sequence 7, Appl
35	11.8	78.7	2454 4 US-09-221-527-7	Sequence 7, Appl
36	11.8	78.7	2454 4 US-09-221-236-7	Sequence 7, Appl
37	11.8	78.7	2454 4 US-09-221-416-7	Sequence 7, Appl
38	11.8	78.7	2454 4 US-09-221-245-7	Sequence 7, Appl
39	11.8	78.7	2454 4 US-09-163-115-7	Sequence 7, Appl
40	11.8	78.7	2454 4 US-09-221-528-7	Sequence 7, Appl
41	11.8	78.7	2454 4 US-09-593-553-7	Sequence 7, Appl
42	11.8	78.7	2687 1 US-08-149-695-8	Sequence 8, Appl
43	11.8	78.7	2687 1 US-08-377-228-8	Sequence 8, Appl
44	11.8	78.7	2873 1 US-08-149-695-1	Sequence 1, Appl
45	11.8	78.7	2873 1 US-08-377-228-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-058-489-14
Sequence 14, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WH197-08PA
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 3408
TYPE: DNA
ORGANISM: Human
US-09-058-489-14

Query Match 89.3%; Score 13.4; DB 4; Length 3408;
Best local similarity 93.3%; Prod. No. 14;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcgggtcattaga 15
DB 728 agcgggtcattaga 742

RESULT 2
US-09-058-489-13
Sequence 13, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WH197-08PA
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 13
LENGTH: 5322
TYPE: DNA
ORGANISM: Human
US-09-058-489-13

Query Match 89.3%; Score 13.4; DB 4; Length 5322;
Best Local Similarity 93.3%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 accgggctattaga 15
|||||
Db 728 accgggctattaga 742

RESULT 3
US-08-981-803-1
Sequence 1, Application US/08981803
Patent No. 6147279
GENERAL INFORMATION:
APPLICANT: POULSEN, PETER
TITLE OF INVENTION: INHIBITION OF GENE EXPRESSION
FILE REFERENCE: 674509-2004
CURRENT APPLICATION NUMBER: US/08/981,803
EARLIER FILING DATE: 1997-04-17
EARLIER APPLICATION NUMBER: PCT/EP96/03052
EARLIER FILING DATE: 1996-07-12
EARLIER APPLICATION NUMBER: 9514435.8
EARLIER FILING DATE: 1995-07-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1167
TYPE: DNA
ORGANISM: Solanum tuberosum
US-08-981-803-1

Query Match 82.7%; Score 12.4; DB 4; Length 1167;
Best Local Similarity 92.9%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ggggctctattaga 15
|||||
Db 537 ggggctctattaga 550

RESULT 4
US-08-981-803-15/C
Sequence 15, Application US/08981803
Patent No. 6147279
GENERAL INFORMATION:
APPLICANT: POULSEN, PETER
TITLE OF INVENTION: INHIBITION OF GENE EXPRESSION
FILE REFERENCE: 674509-2004
CURRENT APPLICATION NUMBER: US/08/981,803
EARLIER FILING DATE: 1997-04-17
EARLIER APPLICATION NUMBER: PCT/EP96/03052
EARLIER FILING DATE: 1996-07-12
EARLIER APPLICATION NUMBER: 9514435.8
EARLIER FILING DATE: 1995-07-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 1167
TYPE: DNA
ORGANISM: Solanum tuberosum
US-08-981-803-15

Query Match 82.7%; Score 12.4; DB 4; Length 1167;
Best Local Similarity 92.9%; Pred. No. 52;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 ggggctctattaga 15
|||||
Db 631 GAGGCTCTATTAGA 618

RESULT 5
US-08-505-486-96/C
Sequence 96, Application US/08505486
Patent No. 5955573
GENERAL INFORMATION:
APPLICANT: Jesse M. Jaynes
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROTIMEL, FIGG, ERNST & KURZ
STREET: 555 Thirtieth Street N.W.
CITY: Washington
STATE: D. C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
COMPUTER: IBM COMPATIBLE
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/505,486
FILING DATE: 21-JUL-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/279,472
FILING DATE: 22-JUL-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: WALKER, BARBARA W.
REGISTRATION NUMBER: 35,400
REFERENCE/DOCKET NUMBER: 2093-117A
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 2022
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE STRANDED
TOPOLOGY: LINEAR
MOLECULAR TYPE:
DESCRIPTION: CHROMIC DNA
US-08-505-486-96

Query Match 82.7%; Score 12.4; DB 2; Length 2022;
Best Local Similarity 92.9%; Pred. No. 55;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ggggctctattaga 15
|||||
Db 647 GAGGCTCTATTAGA 634

RESULT 6
US-08-801-028-96/C
Sequence 96, Application US/08801028
Patent No. 6018102
GENERAL INFORMATION:
APPLICANT: JOAN GARBARTINO
APPLICANT: JESSE M. JAYNES
APPLICANT: WILLIAM BELKNAP
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE CONSTRUCTS, PROTEIN

```

? COUNTRY: USA
? ZIP: 20004
? COMPUTER READABLE FORM:
? MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
? COMPUTER: IBM COMPATIBLE
? OPERATING SYSTEM: DOS
? SOFTWARE: Wordperfect 5.1+
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/340,154
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/505,486
? FILING DATE: 21-JUL-1995
? APPLICATION NUMBER: U.S. 08/279,472
? FILING DATE: 22-JUL-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: WALKER, BARBARA W.
? REGISTRATION NUMBER: 35,400
? REFERENCE/DOCKET NUMBER: 2093-117A
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202)783-6040
? TELEFAX: (202)783-6031
? INFORMATION FOR SEQ ID NO: 96:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2022
? TYPE: NUCLEIC ACID
? STRANDEDNESS: DOUBLE STRANDED
? TOPOLOGY: LINEAR
? MOLECULE TYPE:
? DESCRIPTION: GENOMIC DNA
US-09-340-154-96

Query Match 82.7%; Score 12.4; DB 3; Length 2022;
Best local Similarity 92.9%; Pred. No. 55;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY 2 ggcggcctattaga 15
1 | |||||
Db 647 GAGGGCTATTAGA 634

RESULT 8
PCT-US95-09338-96/C
; Sequence 96, Application PC/TUS9509338
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: URIDUPTIN-1-UTIC PEPTIDE FUSION GENE
; TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
; NUMBER OF SEQUENCES: 98
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1+
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09338
; FILING DATE: 21-JUL-1994
; PRIOR APPLICATION DATA: 08/279,472
; APPLICATION NUMBER: 08/279,472
; FILING DATE: 22-JUL-1994
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2022
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE STRANDED
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; DESCRIPTION: GENOMIC DNA
PCT-US95-09338-96

```

Query Match:	82.7%	Score 12.4:	DB 5:	Length 2022;
Host Local Similarity:	92.9%	Pred. No. 55:		
Matches 13:	Conservative	0:	Mismatches 1:	Indels 0:
Gaps				
QY	2	gggggctcattaga	15	
Db	647	gAGGGCTCAATTAGA	634	

```

1      RESULT 9
2      PCT-US95-09339-96/c
3      : Sequence 96, Application PC/TUS9509339
4      : GENERAL INFORMATION:
5      : APPLICANT:
6      : APPLICANT:
7      : TITLE OF INVENTION: UNIOUTIN-1,LYTIC PEPTIDE FUSION GENE
8      : TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
9      : TITLE OF INVENTION: METHODS OF MAKING AND USING THE SAME
10     : NUMBER OF SEQUENCES: 98
11     : COMPUTER READABLE FORM:
12     : MEDIUM TYPE: Floppy disk
13     : COMPUTER: IBM PC compatible
14     : OPERATING SYSTEM: PC-DOS/MS-DOS
15     : SOFTWARE: WORDPERFECT 5.1+
16     : CURRENT APPLICATION DATA:
17     : APPLICATION NUMBER: PCT/US95/09339
18     : FILING DATE: 21-JUL-1994
19     : PRIOR APPLICATION DATA:
20     : APPLICATION NUMBER: 08/279,472
21     : FILING DATE: 22-JUL-1994
22     : INFORMATION FOR SEQ ID NO: 96:
23     : SEQUENCE CHARACTERISTICS:
24     : LENGTH: 2022
25     : TYPE: NUCLEIC ACID
26     : STRANDEDNESS: DOUBLE STRANDED
27     : TOPOLOGY: LINEAR
28     : MOLECULE TYPE:
29     : DESCRIPTION: GENOMIC DNA
30     : PCT-US95-09339-96

```

Query Match	82.7%	Score 12.4	DB 5	Length 2022
Best Local Similarity	92.9%	Pred. No. 55		
Matches 13; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
Cy	2	gcgcggtcctataga	15	
	1			
	647	GAAGGCTCAATTAGA	634	

RESULT 10
 US-08-505-486-95/C
 ; Sequence 95 Application US/08505486
 ; Patent No. 593573
 ; GENERAL INFORMATION:
 ; APPLICANT: Jesse M. Jaynes
 ; TITLE OF INVENTION: UNICUTIN-LYTIC PEPTIDE FUSION GENE
 ; TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
 ; TITLE OF INVENTION: METHODS OF MAKING AND USING SAME
 ; NUMBER OF SEQUENCES: 98
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ROTHWELL, FIGG, ERNST & KURZ
 ; STREET: 555 Thirteenth Street N.W.
 ; CITY: Washington
 ; STATE: D. C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
 ; COMPUTER: IBM COMPATIBLE
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: Wordperfect 5.1+

```

1      CURRENT APPLICATION DATA:
2      APPLICATION NUMBER:  US/08/505,486
3      FILING DATE:  21-JUL-1995
4      CLASSIFICATION:  536
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER:  U.S.  08/279,472
7      FILING DATE:  22-JUL-1994
8      CLASSIFICATION:  536
9      ATTORNEY/AGENT INFORMATION:
10     NAME:  WALKER, BARBARA W.
11     REGISTRATION NUMBER:  35,400
12     REFERENCE/DOCKET NUMBER:  2093-117A
13     TELECOMMUNICATION INFORMATION:
14     TELEPHONE:  (202)783-6040
15     TELEFAX:  (202)783-6031
16     INFORMATION FOR SHO, ID NO:  95:
17     SEQUENCE CHARACTERISTICS:
18     LENGTH:  2127
19     TYPE:  NUCLEIC ACID
20     STRANDEDNESS:  DOUBLE STRANDED
21     TOPOLOGY:  LINEAR
22     MOLECULE TYPE:
23     DESCRIPTION:  GENOMIC DNA AND OTHER DNA
24
25  US-08-505-486-95

```

```

Query Match1:      82.7%;   Score 12.4;   DB 2;   length 2127;
Best Local Similarity 92.9%;   Prod. No. 15;
Matches 13;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;

QY      2 gcgcgcgtcattaga 15
        | | | | | | | | | |
Db      647 GAGCGCTCTATTAGA 634

```

RESULT 11
 US-08-801-028-95/c
 : Sequence 95, Application US/08801028
 Patent No. 6018102
 GENERAL INFORMATION:
 APPLICANT: JOAN GARBARINO
 APPLICANT: JESSE M. JAYNES
 APPLICANT: WILLIAM BELKNAP
 TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE CONSTRUCTS, PROTEIN
 NUMBER OF SEQUENCES: 98
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: STEVEN J. HUIFQUIST
 ADDRESSEE: INTELLECTUAL PROPERTY/TECHNOLOGY LAW
 STREET: 200 PARK DRIVE, SUITE 210
 STREET: P.O. BOX 14329
 CITY: RESEARCH TRIANGLE PARK
 STATE: NORTH CAROLINA
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
 COMPUTER: APPLE MACINTOSH
 OPERATING SYSTEM: MACINTOSH
 SOFTWARE: M.S. WORD 5.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/801,028
 FILING DATE: 19-FEB-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/279,472
 FILING DATE: JULY 22, 1994
 APPLICATION NUMBER: 08/225,476
 FILING DATE: 04-20-94
 APPLICATION NUMBER: 08/225,476
 FILING DATE: 04-08-94
 APPLICATION NUMBER: 08/039,620
 FILING DATE: 06-04-93
 APPLICATION NUMBER: 08/148,491
 FILING DATE: 11-08-93

APPLICATION NUMBER: 08/148,889
FILING DATE: 11-08-93
ATTORNEY/AGENT INFORMATION:
NAME: WASSERMAN, FRANK S.
REGISTRATION NUMBER: 34,273
REFERENCE/DOCKET NUMBER: 4013-104
TELEPHONE: (919)990-9531
TELEFAX: (919)990-9532
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 2127
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE STRANDED
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: GENOMIC DNA AND OTHER DNA
US-08-801-028-95

Query Match 82.7%; Score 12.4; DB 3; Length 2127;
Best Local Similarity 92.9%; Pred. No. 55;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ggcggctctattaga 15
| |||||
DB 647 GAGGCTCTATTAGA 634

RESULT 12
US-09-340-154-95/c
Sequence 95, Application US/09340154
Patent No. 6084156
GENERAL INFORMATION:
APPLICANT: Jesse M. Jaynes
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
TITLE OF INVENTION: METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESS: ROTHWELL, FIGG, KRIST & KURZ
STREET: 555 Thirteenth Street N.W.
CITY: Washington
STATE: D. C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
COMPUTER: IBM COMPATIBLE
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1+
CURRENT APPLICATION DATA: 09/340.154
APPLICATION NUMBER: US/09/340.154
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/505,486
FILING DATE: 21-JUL-1995
APPLICATION NUMBER: U.S. 08/279,472
FILING DATE: 22-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: WALKER, BARBARA W.
REGISTRATION NUMBER: 35,400
REFERENCE/DOCKET NUMBER: 2093-117A
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 2127
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE STRANDED
TOPOLOGY: LINEAR

MOLECULE TYPE:
DESCRIPTION: GENOMIC DNA AND OTHER DNA
US-09-340-154-95

Query Match 82.7%; Score 12.4; DB 3; Length 2127;
Best Local Similarity 92.9%; Pred. No. 55;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ggcggctctattaga 15
| |||||
DB 647 GAGGCTCTATTAGA 634

RESULT 13
PCT-US95-09338-95/c
Sequence 95, Application PCT/US9509338
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
TITLE OF INVENTION: METHODS OF MAKING AND USING THE SAME
NUMBER OF SEQUENCES: 98
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09338
FILING DATE: 21-JUL-1994
PRIOR APPLICATION DATA: 08/279,472
APPLICATION NUMBER: 08/279,472
FILING DATE: 22-JUL-1994
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 2127
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE STRANDED
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: GENOMIC DNA AND OTHER DNA
PCT-US95-09338-95

Query Match 82.7%; Score 12.4; DB 5; Length 2127;
Best Local Similarity 92.9%; Pred. No. 55;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ggcggctctattaga 15
| |||||
DB 647 GAGGCTCTATTAGA 634

RESULT 14
PCT-US95-09339-95/c
Sequence 95, Application PCT/US9509339
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
TITLE OF INVENTION: METHODS OF MAKING AND USING THE SAME
NUMBER OF SEQUENCES: 98
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09339
FILING DATE: 21-JUL-1994
PRIOR APPLICATION DATA:

117: qb_est148:*
 118: qb_est149:*
 119: qb_est150:*
 120: qb_est151:*
 121: qb_est152:*
 122: qb_est153:*
 123: qb_est154:*
 124: qb_est155:*
 125: qb_est156:*
 126: qb_est157:*
 127: qb_est158:*
 128: qb_est159:*
 129: qb_est160:*
 130: qb_est161:*
 131: qb_est162:*
 132: qb_est163:*
 133: qb_est164:*
 134: qb_est165:*
 135: qb_est166:*
 136: qb_est167:*
 137: qb_est168:*
 138: qb_est169:*
 139: qb_est170:*
 140: qb_est171:*
 141: qb_est172:*
 142: qb_est173:*
 143: qb_est174:*
 144: qb_est175:*
 145: qb_est176:*
 146: qb_est177:*
 147: qb_est178:*
 148: qb_est179:*
 149: qb_est180:*
 150: qb_est181:*
 151: qb_est182:*
 152: qb_est183:*
 153: qb_est184:*
 154: qb_est185:*
 155: qb_est186:*
 156: qb_est187:*
 157: qb_est188:*
 158: qb_est189:*
 159: qb_est190:*
 160: qb_est191:*
 161: qb_est192:*
 162: qb_est193:*
 163: qb_est194:*
 164: qb_est195:*
 165: qb_est196:*
 166: qb_est197:*
 167: qb_est198:*
 168: qb_est199:*
 169: qb_est200:*
 170: qb_est201:*
 171: qb_est202:*
 172: qb_est203:*
 173: qb_est204:*
 174: qb_est205:*
 175: qb_est206:*
 176: qb_est207:*
 177: qb_est208:*
 178: qb_est209:*
 179: qb_est210:*
 180: qb_est211:*
 181: qb_est212:*
 182: qb_est213:*
 183: qb_est214:*
 184: qb_est215:*
 185: qb_est216:*
 186: qb_est217:*
 187: qb_est218:*
 188: qb_est219:*
 189: qb_est220:*

190: em_gss_pln1:*
 191: em_gss_pln2:*
 192: em_gss_pro:*
 193: em_gss_rtd1:*
 194: em_gss_rtd2:*
 195: em_gss_rtd3:*
 196: em_gss_rtd4:*
 197: em_gss_rtd5:*
 198: em_gss_vrt1:*
 199: em_gss_vrt2:*
 200: em_gss_vrt3:*
 201: qb_gss1:*
 202: qb_gss2:*
 203: qb_gss3:*
 204: qb_gss4:*
 205: qb_gss5:*
 206: qb_gss6:*
 207: qb_gss7:*
 208: qb_gss8:*
 209: qb_gss9:*
 210: qb_gss10:*
 211: qb_gss11:*
 212: qb_gss12:*
 213: qb_gss13:*
 214: qb_gss14:*
 215: qb_gss15:*
 216: qb_gss16:*
 217: qb_gss17:*
 218: qb_gss18:*
 219: qb_gss19:*
 220: qb_gss20:*
 221: qb_gss21:*
 222: qb_gss22:*
 223: qb_gss23:*
 224: qb_gss24:*
 225: qb_gss25:*
 226: qb_gss26:*
 227: qb_gss27:*
 228: qb_gss28:*
 229: qb_gss29:*
 230: qb_gss30:*
 231: qb_gss31:*
 232: qb_gss32:*
 233: qb_gss33:*
 234: qb_gss34:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	446	211	A0774045	A0774045 HS-2190.B
2	14	93.3	171	226	H000027	B000027 CSRL-100E3
3	14	93.3	412	203	A0166430	A0166430 HS-3128.B
4	14	93.3	570	167	BE436368	BE436368 EST407446
5	14	93.3	616	144	BF113023	BF113023 EST440533
6	13.4	89.3	146	119	AW723826	AW723826 FI1050m.f
7	13.4	89.3	149	18	A1320373	A1320373 C3H050m.f
8	13.4	89.3	151	119	AW724282	AW724282 f4906nm.f
9	13.4	89.3	151	119	AW724283	AW724283 f4906nm.f
10	13.4	89.3	187	119	AW715484	AW715484 g5c07nm.f
11	13.4	89.3	189	18	A1321118	A1321118 d4d10nm.f
12	13.4	89.3	201	133	BA429584	BA429584 B429584
13	13.4	89.3	308	127	BB200417	BB200417 BB200417
14	13.4	89.3	316	201	A0072101	A0072101 HS-3010.B
15	13.4	89.3	323	18	A1318822	A1318822 a2103nm.f
16	13.4	89.3	326	111	AW171957	AW171957 618056E01
17	13.4	89.3	368	218	A2267638	A2267638 RPCI-23-1
18	13.4	89.3	391	1	AA003631	AA003631 mg60C05.r

LOCUS				
AQ166430	412 bp	DNA		
		GSS		
HS_3128-B1_G03_MR	CIT	Approved	Human	Genomic Sperm Library D Homolog
				16-OCT-1998

ACCESSION	ASapiens genomic clone Plate-3128 Col-5 Row-N, DNA sequence.		
VERSION	A0166430.1 GI:3564625		
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
AUTHORS	1 (bases 1 to 412) Kellers, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Hood, L., Shaker, R., Hurlong, J., Young, O., Zhao, S., Adams, M.D. and		
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome		
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)		
REMARK	99380589		
COMMENT	Contact: Mahairas CG, Wallace JC, Hood L, High Throughput Sequencing Center University of Washington Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallaceu.washington.edu Sequence Tagged Connector Plate: 3128 row: N column: 5 Class: BAC ends High quality sequence stop: 412. Location/Qualifiers 1..412 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="Plate-3128 Col-5 Row-N" /clone_lib="C11 Approved Human Genomic Sperm Library B" /sex="male" /note="Organ: sperm; Vector: pBeloBAC11; BAC clones in E-coil DH10B"		
BASE COUNT	108 a 84 c 77 g 143 t		
ORIGIN			
Query Match	93.3%; Score 14; DB 203; length 412;		
Best Local Similarity	100.0%; Pred. No. 1.9e+02;		
Matches	14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 accgagctatcatg 14 		
DB	356 ACCCGCTCTATTAC 369		
RESULT	4		
LOCUS	BE436368 570 bp mRNA HST 24-JUL-2000		
DEFINITION	EST407446 tomato breaker fruit, T1CR lycopersicon esculentum cDNA clone CDEG32M3, mRNA sequence.		
ACCESSION	BE436368		
VERSION	BE436368.1 GI:9434211		
KEYWORDS	EST.		
SOURCE	tomato.		
ORGANISM	Lycopersicon esculentum		
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.		
AUTHORS	Alcala, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, L.E., Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Kohnig, G.M., Nieman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.		
TITLE	Generation of ESTs from tomato fruit tissue, breaker stage		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: David Frisch Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA		

Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfritsch@clemson.edu
 5 prime sequence.
 location/Qualifiers
 1..570
 /organism="Lycopersicon esculentum"
 /cultivar="M496"
 /db_xref="taxon:4081"
 /clone="cLEG32M3"
 /clone_1db="tomato breaker fruit, TIGR"
 /tissue_type="Pericarp"
 /dev_stage="breaker"
 /lab_host="SOLR"
 /note="Vector: pBluescriptSKmCuaapt; Site_1: EcoRI;
 Site_2: XhoI; Fruit were harvested at the breaker stage
 (first sign of lycopene accumulation on the blossom end o
 the fruit). Fruit were cut in half and the seeds and
 locules were discarded prior to freezing the pericarp."
 BASE COUNT 186 a 128 c 83 g 173 t
 ORIGIN
 Query Match 93.3%; Score 14; DB 167; Length 570;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0
 Oy 1 aacggagctctatag 14
 |||||||||
 Db 241 AGCGGCTATTAG 254
 RESULT 5
 BP113023 616 bp mRNA EST 20-Oct-2000
 LOCUS EST440533 tomato breaker fruit Lycopersicon esculentum cDNA clone
 DEFINITION cLEG42L2 5' sequence, mRNA sequence.
 ACCESSION BP113023
 VERSION BP113023.1 GI:10942633
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta
 Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasteridII
 1 (bases 1 to 616)
 REFERENCE
 AUTHORS Alcala,J., Vrechalov,J., White,K., Van der Hoeven,R.S., Holt,I.E.,
 Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Konig,C.M.,
 Niernau,W., Fraser,C.M., Martin,C.B., Giovannoni,J.J. and Tanksley
 ,S.D.
 TITLE Generation of ESTs from tomato fruit tissue, breaker stage
 JOURNAL Unpublished (2000)
 COMMENT Contact: David Fritsch
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfritsch@clemson.edu.
 FEATURES
 source
 location/Qualifiers
 1..616
 /organism="Lycopersicon esculentum"
 /cultivar="M496"
 /db_xref="taxon:4081"
 /clone="cLEG42L2"
 /clone_1db="tomato breaker fruit"
 /tissue_type="Pericarp"
 /dev_stage="breaker"
 /lab_host="SOLR"
 /note="Vector: pBluescriptSKmCuaapt; Site_1: EcoRI;
 Site_2: XhoI; Fruit were harvested at the breaker stage
 (first sign of lycopene accumulation on the blossom end of
 the fruit). Fruit were cut in half and the seeds and

locules were discarded prior to freezing the pericarp."

BASE COUNT 199 a 138 c 90 g 189 t

ORIGIN

Query Match 93.3%; Score 14; DB 144; Length 616;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agcggctctatag 14
|||||

Db 245 ACCGGGCTCTATTAG 258

RESULT 6
AW723826 146 bp mRNA EST 19-APR-2000
LOCUS F1105nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION
AW723826
ACCESSION
AW723826
VERSION
AW723826.1 GI:7618386
KEYWORDS
EST.
SOURCE
Neurospora crassa.
ORGANISM
Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.

REFERENCE
1 (bases 1 to 146)
Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.
Two Neurospora crassa EST Databases
Unpublished (1998)
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu

COMMENT
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Forward Primer
High quality sequence stop: 121.
Location/Qualifiers
1..146

FEATURES

source
/organism="Neurospora crassa"
/strain="bd, frq7 A"
/db_xref="taxon:5141"
/clone="F1105nm"
/clone_lib="Neurospora crassa morning cDNA library"
/tissue_type="tissue harvested following 22hr growth in dark"
/note="Vector: pBluescript SK-; Site.1: XbaI; Site.2: EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript."
BASE COUNT 58 a 26 c 18 g 44 t

Query Match 89.3%; Score 13.4; DB 119; Length 146;
Best Local Similarity 93.3%; Pred. No. 4.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcggctctatag 15
|||||

Db 49 AGCGGCTCTATTATA 35

RESULT 7
A1320373 149 bp mRNA EST 18-DEC-1998
LOCUS C3h05nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION
CDNA clone c3h05nm 5', mRNA sequence.
ACCESSION
A1320373

VERSION
A1320373.1 GI:4036355
EST.

KEYWORDS

Neurospora crassa.

Neurospora crassa

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Sordariales; Sordariaceae; Neurospora.

1 (bases 1 to 149)

Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.

Two Neurospora crassa EST Databases

Unpublished (1998)

Other ESTs: C3h05nm.f1

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal

Genetics Stock Center

Seq primer: Universal Forward Primer

High quality sequence stop: 127.

Location/Qualifiers

1..149

/organism="Neurospora crassa"

/strain="bd, frq7 A"

/db_xref="taxon:5141"

/clone="C3h05nm"

/clone_lib="Neurospora crassa morning cDNA library"

/tissue_type="tissue harvested following 22hr growth in dark"

/note="Vector: pBluescript SK-; Site.1: XbaI; Site.2: EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript."

BASE COUNT 59 a 26 c 19 g 45 t

ORIGIN

QY 1 agcggctctatag 15

|||||

Db 49 AGCGGCTCTATTATA 35

Query Match 89.3%; Score 13.4; DB 18; Length 149;
Best Local Similarity 93.3%; Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 8
AW724282 151 bp mRNA EST 19-APR-2000
LOCUS F1406nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION
CDNA clone F1406nm 5', mRNA sequence.
ACCESSION
AW724282
VERSION
AW724282.1 GI:7618842
KEYWORDS
EST.
SOURCE
Neurospora crassa.
ORGANISM
Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.

1 (bases 1 to 151)

Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.

Two Neurospora crassa EST Databases

Unpublished (1998)

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal

Genetics Stock Center

Seq primer: Universal Forward Primer

High quality sequence stop: 127.

Location/Qualifiers

1..149

/organism="Neurospora crassa"

/strain="bd, frq7 A"

/db_xref="taxon:5141"

/clone="C3h05nm"

/clone_lib="Neurospora crassa morning cDNA library"

/tissue_type="tissue harvested following 22hr growth in dark"

/note="Vector: pBluescript SK-; Site.1: XbaI; Site.2: EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript."

BASE COUNT 59 a 26 c 19 g 45 t

ORIGIN

QY 1 agcggctctatag 15

|||||

Db 49 AGCGGCTCTATTATA 35

Query Match 89.3%; Score 13.4; DB 18; Length 149;
Best Local Similarity 93.3%; Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 8
AW724282 151 bp mRNA EST 19-APR-2000
LOCUS F1406nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION
CDNA clone F1406nm 5', mRNA sequence.
ACCESSION
AW724282
VERSION
AW724282.1 GI:7618842
KEYWORDS
EST.
SOURCE
Neurospora crassa.
ORGANISM
Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.

1 (bases 1 to 151)

Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.

Two Neurospora crassa EST Databases

Unpublished (1998)

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal

Genetics Stock Center

Seq primer: Universal Forward Primer

High quality sequence stop: 127.

Location/Qualifiers

1..149

/organism="Neurospora crassa"

/strain="bd, frq7 A"

/db_xref="taxon:5141"

/clone="C3h05nm"

/clone_lib="Neurospora crassa morning cDNA library"

/tissue_type="tissue harvested following 22hr growth in dark"

/note="Vector: pBluescript SK-; Site.1: XbaI; Site.2: EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript."

BASE COUNT 59 a 26 c 19 g 45 t

ORIGIN

QY 1 agcggctctatag 15

|||||

Db 49 AGCGGCTCTATTATA 35

Query Match 89.3%; Score 13.4; DB 18; Length 149;
Best Local Similarity 93.3%; Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 8
AW724282 151 bp mRNA EST 19-APR-2000
LOCUS F1406nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION
CDNA clone F1406nm 5', mRNA sequence.
ACCESSION
AW724282
VERSION
AW724282.1 GI:7618842
KEYWORDS
EST.
SOURCE
Neurospora crassa.
ORGANISM
Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.

1 (bases 1 to 151)

Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.

Two Neurospora crassa EST Databases

Unpublished (1998)

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal

Genetics Stock Center

Seq primer: Universal Forward Primer

High quality sequence stop: 127.

Location/Qualifiers

1..149

/organism="Neurospora crassa"

/strain="bd, frq7 A"

/db_xref="taxon:5141"

/clone="C3h05nm"

/clone_lib="Neurospora crassa morning cDNA library"

/tissue_type="tissue harvested following 22hr growth in dark"

/note="Vector: pBluescript SK-; Site.1: XbaI; Site.2: EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript."

BASE COUNT 59 a 26 c 19 g 45 t

ORIGIN

QY 1 agcggctctatag 15

|||||

Db 49 AGCGGCTCTATTATA 35

Query Match 89.3%; Score 13.4; DB 18; Length 149;
Best Local Similarity 93.3%; Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 8
AW724282 151 bp mRNA EST 19-APR-2000
LOCUS F1406nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION
CDNA clone F1406nm 5', mRNA sequence.
ACCESSION
AW724282
VERSION
AW724282.1 GI:7618842
KEYWORDS
EST.
SOURCE
Neurospora crassa.
ORGANISM
Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.

1 (bases 1 to 151)

Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.

Two Neurospora crassa EST Databases

Unpublished (1998)

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal

Genetics Stock Center

Seq primer: Universal Forward Primer

High quality sequence stop: 127.

Location/Qualifiers

1..149

/organism="Neurospora crassa"

/strain="bd, frq7 A"

/db_xref="taxon:5141"

/clone="C3h05nm"

/clone_lib="Neurospora crassa morning cDNA library"

/tissue_type="tissue harvested following 22hr growth in dark"

/note="Vector: pBluescript SK-; Site.1: XbaI; Site.2: EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript."

BASE COUNT 59 a 26 c 19 g 45 t

ORIGIN

QY 1 agcggctctatag 15

|||||

Db 49 AGCGGCTCTATTATA 35

Query Match 89.3%; Score 13.4; DB 18; Length 149;
Best Local Similarity 93.3%; Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 8
AW724282 151 bp mRNA EST 19-APR-2000
LOCUS F1406nm.f1 Neurospora crassa morning cDNA library Neurospora crassa
DEFINITION
CDNA clone F1406nm 5', mRNA sequence.
ACCESSION
AW724282
VERSION
AW724282.1 GI:7618842
KEYWORDS
EST.
SOURCE
Neurospora crassa.
ORGANISM
Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.

1 (bases 1 to 151)

Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.

Two Neurospora crassa EST Databases

Unpublished (1998)

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal

Genetics Stock Center

Seq primer: Universal Forward Primer

High quality sequence stop: 127.

Location/Qualifiers

1..149

/organism="Neurospora crassa"

/strain="bd, frq7 A"

/db_xref="taxon:5141"

/clone="C3h05nm"

/clone_lib="Neurospora crassa morning cDNA library"

/tissue_type="tissue harvested following 22hr growth in dark"

/note="Vector: pBluescript SK-; Site.1: XbaI; Site.2: EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript."

Seq primer: Universal Forward Primer
High quality sequence stop: 116.
Location/Qualifiers
1. 151

source

/organism="Neurospora crassa"
/strain="bd, frq7 A"
/db_xref="taxon:5141"
/clone="f4g06nm"
/clone_lib="Neurospora crassa morning cDNA library"
/tissue_type="tissue harvested following 22hr growth in dark"

/note="Vector: pBluescript SK-; Site_1: XbaI; Site_2: EcoRI; See: Bell-Pedersen, D., et al. PNAS 93:13096,1996. 5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript"

BASE COUNT
ORIGIN

59 a 26 c 21 g 45 t

Query Match

Best Local Similarity 89.3%; Score 13.4; DB 119; Length 151;
Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agcgggtctattaga 15
|||||
Db 49 AGCGGCTCTATTATA 35

RMSUT 9
AM724283/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broeou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Forward Primer
High quality sequence stop: 116.
Location/Qualifiers
1. 151

FEATURES

source

/organism="Neurospora crassa"
/strain="bd, frq7 A"
/db_xref="taxon:5141"
/clone="f4g06nm"
/clone_lib="Neurospora crassa morning cDNA library"
/tissue_type="tissue harvested following 22hr growth in dark"

/note="Vector: pBluescript SK-; Site_1: XbaI; Site_2: EcoRI; See: Bell-Pedersen, D., et al. PNAS 93:13096,1996. 5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript"

BASE COUNT

59 a 26 c 21 g 45 t

Query Match

89.3%; Score 13.4; DB 119; Length 151;

Best Local Similarity 93.3%; Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agcgggtctattaga 15
|||||
Db 49 AGCGGCTCTATTATA 35

RESULT 10
AM715484/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broeou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Forward Primer
High quality sequence stop: 152.
Location/Qualifiers
1. 187

/organism="Neurospora crassa"
/strain="bd, frq7 A"
/db_xref="taxon:5141"
/clone="g5c07nm"
/clone_lib="Neurospora crassa morning cDNA library"
/tissue_type="tissue harvested following 22hr growth in dark"

FEATURES

source

/note="Vector: pBluescript SK-; Site_1: XbaI; Site_2: EcoRI; See: Bell-Pedersen, D., et al. PNAS 93:13096,1996. 5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript"

BASE COUNT

69 a 35 c 23 g 60 t

Query Match 89.3%; Score 13.4; DB 119; Length 187;
Best Local Similarity 93.3%; Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agcgggtctattaga 15
|||||
Db 49 AGCGGCTCTATTATA 35

RESULT 11
A1321118/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broeou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Forward Primer
High quality sequence stop: 152.
Location/Qualifiers
1. 187

JOURNAL
COMMENT

Unpublished (1998)
 Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Partington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal
 Genetics Stock Center
 Seq primer: Universal Forward Primer
 High quality sequence stop: 273.

FEATURES

source

1. .323
 location/Qualifiers
 /organism="Neurospora crassa"
 /strain="pd, lrg A"
 /db_xref="taxon:5141"
 /clone="a2f03nm"
 /clone_lib="Neurospora crassa morning cDNA library"
 /tissue_type="tissue harvested following 22hr growth in
 dark"
 /note="Vector: pBluescript SK-; site_1: XbaI; site_2:
 EcoRI; See: Bell-Pedersen, D., et al. PNAS 93:13096, 1996.
 5' end of cDNA cloned into XbaI site of pBluescript; 3'
 end of cDNA cloned into KcoRI site of pBluescript"

BASE COUNT

111 a 59 c 42 g 111 t

ORIGIN

Query Match

Best Local Similarity 89.3%; Score 13.4; DB 18; Length 323;
 Matches 14; Conservatve 0; Mismatches 1; Indels 0;

QY 1 aacgggtctataga 15
 |||

DB 49 AGCGGCTCTATTATA 35

Search completed: May 1, 2001, 17:25:39
 Job time: 8725 sec